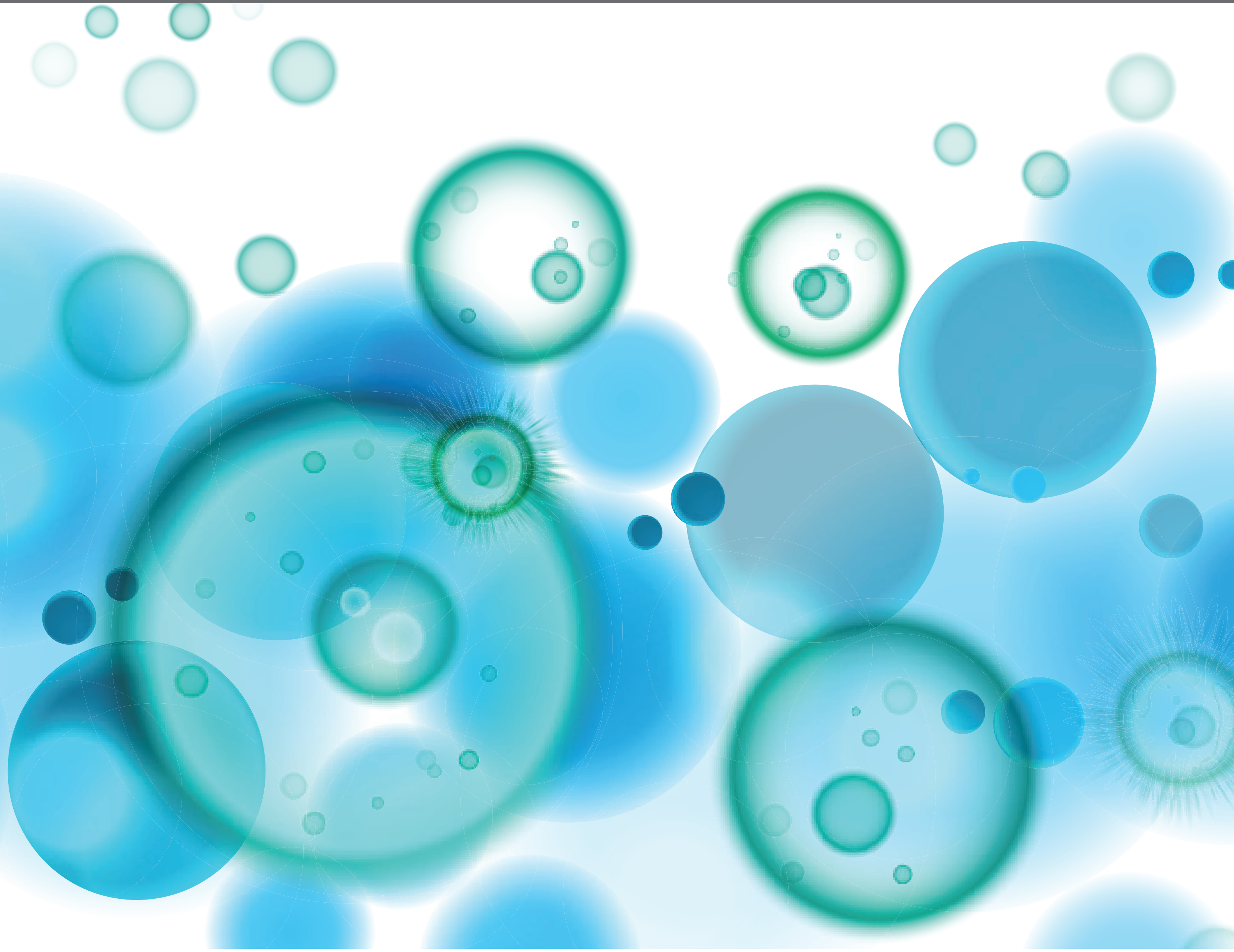
TOWARDS PRECISION MEDICINE FOR IMMUNE-MEDIATED DISORDERS: ADVANCES IN USING BIG DATA AND ARTIFICIAL INTELLIGENCE TO UNDERSTAND HETEROGENEITY IN INFLAMMATORY RESPONSES

EDITED BY: Xu-jie Zhou, Lam Cheung Tsoi and Amanda S. MacLeod
PUBLISHED IN: Frontiers in Immunology and Frontiers in Genetics



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-583-6
DOI 10.3389/978-2-88976-583-6

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

TOWARDS PRECISION MEDICINE FOR IMMUNE-MEDIATED DISORDERS: ADVANCES IN USING BIG DATA AND ARTIFICIAL INTELLIGENCE TO UNDERSTAND HETEROGENEITY IN INFLAMMATORY RESPONSES

Topic Editors:
Xu-jie Zhou, First Hospital, Peking University, China
Lam Cheung Tsoi, University of Michigan, United States
Amanda S. MacLeod, Janssen Research and Development (United States), United States

Topic Editor Dr. MacLeod is employed by Janssen. All other Topic Editors declare no competing interests with regards to the Research Topic subject.

Citation: Zhou, X.-J., Tsoi, L. C., MacLeod, A. S., eds. (2022). Towards Precision Medicine for Immune-Mediated Disorders: Advances in Using Big Data and Artificial Intelligence to Understand Heterogeneity in Inflammatory Responses. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-583-6

Table of Contents

EDITORIAL
published: 22 June 2022
doi: 10.3389/fimmu.2022.948885

Editorial: Advances in Using Big Data and Artificial Intelligence to Understand Heterogeneity in Inflammatory Responses

Xu-jie Zhou[1*†], *Amanda S. MacLeod*[2,3,4*†] and *Lam Cheung Tsoi*[5,6,7*†]

[1] Renal Division, Peking University First Hospital; Kidney Genetic Center, Peking University Institute of Nephrology; Key Laboratory of Renal Disease, Ministry of Health of China, Beijing, China, [2] Department of Dermatology; Molecular Genetics and Microbiology, Duke University, Durham, NC, United States, [3] Department of Immunology; Molecular Genetics and Microbiology, Duke University, Durham, NC, United States, [4] Janssen Research and Development, Spring House, PA, United States, [5] Department of Dermatology, University of Michigan, Ann Arbor, MI, United States, [6] Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI, United States, [7] Department of Biostatistics, University of Michigan, Ann Arbor, MI, United States

OPEN ACCESS

Edited and reviewed by:
Ryu Watanabe,
Osaka Metropolitan University, Japan

****Correspondence:***
Xu-jie Zhou
zhouxujie@bjmu.edu.cn
Amanda S. MacLeod
amanda.macleod@duke.edu
Lam Cheung Tsoi
alextsoi@med.umich.edu

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Autoimmune and Autoinflammatory Disorders, a section of the journal Frontiers in Immunology

Received: *20 May 2022*
Accepted: *06 June 2022*
Published: *22 June 2022*

Citation:
Zhou X-j, MacLeod AS and Tsoi LC (2022) Editorial: Advances in Using Big Data and Artificial Intelligence to Understand Heterogeneity in Inflammatory Responses. Front. Immunol. 13:948885. doi: 10.3389/fimmu.2022.948885

Keywords: artificial intelligence, big data, heterogeneity, immune-mediated disorders, multi-omics, precision medicine

Editorial on the Research Topic

Towards Precision Medicine for Immune-Mediated Disorders: Advances in Using Big Data and Artificial Intelligence to Understand Heterogeneity in Inflammatory Responses

In this Research Topic, we have hosted 3 in-depth reviews and 15 original research articles presenting how novel technological, methodological, and conceptual advancements can be integrated to study the underlying mechanisms that drive the heterogeneity in inflammatory responses among patients suffering from immune-mediated conditions.

The immune system plays a vital role in health and disease, and is regulated through a complex interactive network of immune cells and mediators, thus multi-omics approach in immunological research is advocated to provide a better understanding of the system. As biomedical research transitioning into data-rich science, an era of "big data" emerged owing to these advancements. The integration of such multi- layered datasets with longitudinal assessments of patient outcomes has the capacity to shed important lights into different aspects of disease pathogenesis, progression and cell-specific responses, with which to guide design of targeted therapies. Multi-source big data is thus suggested to be the major driver of precision medicine. However, only data alone can be hardly to be transformed into clinically actionable knowledge, if we don't have proper analysis methods. Thanks to the advances of computing science, artificial intelligence (AI) is developed for robust data analysis.

Chu et al. extensively introduced multi-omics approaches in immunological research, and Orrù et al. reviewed that systematic multi-parametric flow cytometry coupled with high-resolution genetics and transcriptomics can be used to reveal endophenotypes of autoimmune diseases for therapeutic development. Through AI-based analysis of different disease parameters – including clinical and para-clinical outcomes, and molecular profilesl from multi-omic data, a digital twin paired to the patient's characteristic can be created, enabling healthcare professionals to handle large amounts of patient data, and Voigt et al. discussed the use of digital twins for MS as a revolutionary tool to improve diagnosis, monitoring and therapy refining. Digital twins will help make precision medicine and patient-centered care a reality in everyday life. At the level of genomics, through genome-wide association study (GWAS), Connell et al. found a genome-wide significant association between intergenic variant rs35569429 and response to ustekinumab for the treatment of moderate to severe psoriasis. These work also discuss how AI and multi-omics can be applied and integrated, to offer opportunities to develop novel diagnostic and therapeutic means in immune related diseases.

Using transcriptomic analysis on skin biopsies, Abernathy-Close et al. observed that skin-associated B cell responses distinguish discoid lupus erythematosus (DLE) from subacute cutaneous lupus erythematosus (SCLE) and acute cutaneous lupus erythematosus (ACLE). This data has important implications for trial design for patients with isolated cutaneous lupus erythematosus (CLE). Maruyama et al. conduced RNA-seq data analysis and identified several lncRNAs such as *MALAT1, CA3-AS1, GASAL1, PSMA3-AS1, MIR4435-2HG, IL21-AS1, AC111000.4*, and *LINC01501*, and some of them are associated with active *Visceral leishmaniasis* infection. By implementing the weighted gene co-expression network analysis (WGCNA), Zhang et al. suggested that the osteoarticular involvement in psoriasis and ankylosing spondylitis (AS) could be mediated by the mRNA surveillance pathway. Also based on RNA-Seq expression, Cao et al. observed that *RIMKLB* expression is associated with survival outcomes and tumor-infiltrating immune cells (TIICs) in patients with colorectal cancer (CRC), indicating that it might be a potential novel prognostic biomarker that reflects the immune infiltration status. There are also different studies in our Research Topics that utilize single cell genomics approaches. Xu et al. performed single-cell RNA sequencing, demonstrating cell-specific transcriptional profiles in the kidney, anti- phospholipase A2 receptor (*PLA2R*) positive membranous nephropathy (MN) -associated novel genes, signaling pathways involved, and potential pathogenesis concerning ligand-receptor interactions. Liu et al. took single-cell RNA-sequencing of CD45+cells isolated from active lesions of patients with psoriasis vulgaris, they found *CXCL13* significantly correlated with the severity of lesions and genes elevated in psoriatic skin-resident memory T cells are enriched for programs orchestrating chromatin and CDC42-dependent cytoskeleton remodeling. Alber et al. used single cell CITE-Seq (Cellular Indexing of Transcriptomes and Epitopes by sequencing) technology to analyze peripheral blood mononuclear cells (PBMCs) in ankylosing spondylitis (AS) and identified a number of molecular features which were associated with AS were linked with inflammation and other immune-mediated diseases. With the increasing resources of single-cell sequencing data, issue of heterogeneity and limited comprehension of chronic autoimmune disease pathophysiology could be better addressed. Ma et al. integrated several sets of single-cell RNA sequencing data and bulk RNA-sequencing data from open access database deposited in the Gene Expression Omnibus (GEO), and found that the interactions among the peripheral blood mononuclear cells (PBMCs) subpopulations of SLE patients may be weakened under the inflammatory microenvironment. With transcriptomic datasets in ulcerative colitis, Chen et al. applied artificial neural network (ANN) and identified a predictive RNA model in which combination of *CDX2, CHP2, HSD11B2, RANK, NOX4*, and *VDR* was a good predictor of patients' response to infliximab (IFX) therapy. Liu et al. performed single cell profiling of transcriptome and cell surface protein expression to compare the peripheral blood immunocyte populations of individuals with psoriatic arthritis (PSA), individuals with cutaneous psoriasis (PSO) alone, and healthy individuals. They observed a higher abundance of Tregs and dnT cells in PSA patients and a higher abundance of hematopoietic stem precursor cells (HSPCs) in healthy subjects.

O'Neil1 et al. sought to identify serum proteomic alterations that dictate clinically important features of stable rheumatoid

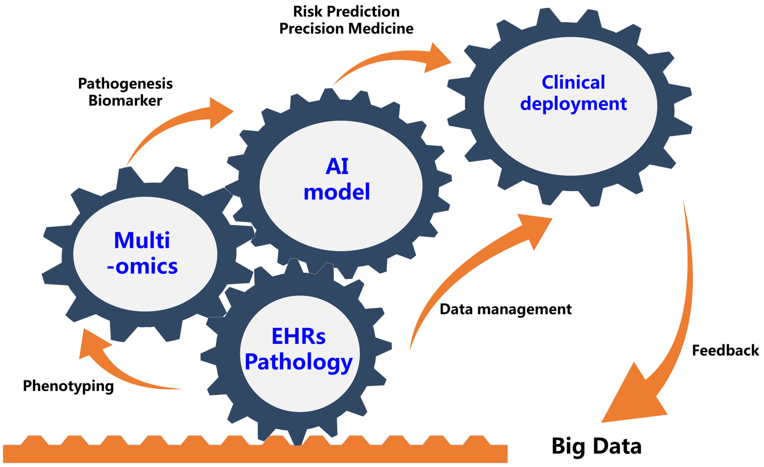


FIGURE 1 | Future Integration of Artificial Intelligence And Multi-omics Will Benefit Precision Medicine for Immune-Mediated Disorders. EHRs, electronic health records.

arthritis, and couple broad-based proteomics with machine learning to predict future flare. They defined 4 proteomic clusters reflecting biological mechanisms, and found an XGboost machine learning algorithm could classify patients who relapsed with an AUC of 0.80 using only baseline serum proteomics. We can also see that some novel data or perspectives were discussed and shared from different groups. For example, Cao et al. took two-sample bidirectional Mendelian randomization analysis and cross-trait meta-analysis between major depressive disorder (MDD) and atopic diseases (AD: asthma, hay fever, and eczema). They found a significant genetic correlation between MDD and ADs, and detected a major causal effect of genetic liability to depression on ADs. Jamerson et al. explored the heterogeneity of atopic dermatitis and psoriasis between African American and European American patients by summarizing epidemiological studies, addressing potential molecular and environmental factors, with a focus on psychosocial or psychological stress on immune pathways, and highlighted the role of the hypothalamus-pituitary-adrenal (HPA) axis and *IL-18* in atopic dermatitis, corticotropin-releasing hormone and brain derived neurotrophic factor in psoriasis, and cortisol levels in both. It supports environmental components in disease heterogeneity and their influence on disease pathogenesis. Observational studies may also shed some light on precision medicine. By retrospectively reviewing the Taiwan National Health Insurance Research Database (NHIRD) within 13 years, Li et al. observed that influenza vaccination was associated with lower asthma risk in patients with AD.

The use of AI and multi-omics in human diseases are still in their infancy, mostly for research. Although it is premature to try and define the potential clinical utility of these newly found molecule biomarkers or predictive models, they do provide an important impetus for further studies that aim to further define a biological definition of sub-phenotypes in patients with immune related diseases that can ultimately guide clinical decision making (**Figure 1**).

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was supported by Beijing Natural Science Foundation (Z190023).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Immunology

REVIEW
published: 03 May 2021
doi: 10.3389/fimmu.2021.669811

Digital Twins for Multiple Sclerosis

*Isabel Voigt, Hernan Inojosa, Anja Dillenseger, Rocco Haase, Katja Akgün and Tjalf Ziemssen**

Center of Clinical Neuroscience, Department of Neurology, University Hospital Carl Gustav Carus, Technical University of Dresden, Dresden, Germany

An individualized innovative disease management is of great importance for people with multiple sclerosis (pwMS) to cope with the complexity of this chronic, multidimensional disease. However, an individual state of the art strategy, with precise adjustment to the patient's characteristics, is still far from being part of the everyday care of pwMS. The development of digital twins could decisively advance the necessary implementation of an individualized innovative management of MS. Through artificial intelligence-based analysis of several disease parameters – including clinical and para-clinical outcomes, multi-omics, biomarkers, patient-related data, information about the patient's life circumstances and plans, and medical procedures – a digital twin paired to the patient's characteristic can be created, enabling healthcare professionals to handle large amounts of patient data. This can contribute to a more personalized and effective care by integrating data from multiple sources in a standardized manner, implementing individualized clinical pathways, supporting physician-patient communication and facilitating a shared decision-making. With a clear display of pre-analyzed patient data on a dashboard, patient participation and individualized clinical decisions as well as the prediction of disease progression and treatment simulation could become possible. In this review, we focus on the advantages, challenges and practical aspects of digital twins in the management of MS. We discuss the use of digital twins for MS as a revolutionary tool to improve diagnosis, monitoring and therapy refining patients' well-being, saving economic costs, and enabling prevention of disease progression. Digital twins will help make precision medicine and patient-centered care a reality in everyday life.

Keywords: multiple sclerosis, precision medicine, personalized medicine, digital twin, decision analysis, medical care

OPEN ACCESS

Edited by:
Xu-jie Zhou,
Peking University First Hospital, China

Reviewed by:
Gleb Radchenko,
South Ural State University, Russia
Koen R. C. Bruynseels,
Delft University of Technology, Netherlands

****Correspondence:***
Tjalf Ziemssen
tjalf.ziemssen@uniklinikum-dresden.de

Specialty section:
This article was submitted to Autoimmune and Autoinflammatory Disorders, a section of the journal Frontiers in Immunology

Received: *19 February 2021*
Accepted: *16 April 2021*
Published: *03 May 2021*

Citation:
Voigt I, Inojosa H, Dillenseger A, Haase R, Akgün K and Ziemssen T (2021) Digital Twins for Multiple Sclerosis. Front. Immunol. 12:669811. doi: 10.3389/fimmu.2021.669811

INTRODUCTION

The technology of digital twins (DTs) is a promising concept that has become the focus of interest in industry and, in recent years, in healthcare sector as well. DTs are a revolutionary tool in phenotyping patients, where analysis of large amounts of data (big data) through new technologies like artificial intelligence (AI) enables visualization of a virtual copy (twin) of the patient at different stages of the disease and supports further therapeutic decisions. However, the use of DTs in medical care and especially in the management of patients is still in its infancy. DTs have enormous potential, especially when it comes to precision medicine: they can be used to simulate individual therapies in advance and visualize potential therapy results and disease progression. The concept of DTs seems to be particularly suitable for the treatment of multiple sclerosis (MS), because this chronic autoimmune "disease of a thousand faces" is characterized by heterogeneous course, complexity and multidimensionality, an increasing number of treatment options and a resulting wealth of data. DTs can significantly improve precision medicine for people with MS (pwMS) by enabling healthcare professionals (HCPs) to handle big data and provide more personalized and effective care. In this paper, we focus on our vision of how to design a DT for the management of MS. The advantages of DTs for the personalized treatment of individual pwMS are highlighted without ignoring the challenges on its development. With our review, we want to answer the question whether "Digital Twins for Multiple Sclerosis" (DTMS) may serve as a game changer in the management of MS.

MULTIPLE SCLEROSIS REQUIRES PRECISION MEDICINE

Multiple Sclerosis as a Chronic Multidimensional Disease

MS is a chronic autoimmune, degenerative and lifelong disease of the central nervous system (CNS) and the most common cause of neurological disability in young adults. At a pathological level, the infiltration of immune cells into the CNS manifests as localized demyelinating lesions in the white and gray matters of the brain and spinal cord, observed in pathological specimens as well as in magnetic resonance imaging (MRI) sequences (1). In addition, the disease leads to a progressive destruction of myelin layers (demyelination) and progressive axonal injury, loss and neurodegeneration, impairing the function of the CNS in several ways (2, 3).

MS has different clinical disease courses that have been classically described. Around 85-90% of the patients are diagnosed with a relapsing remitting form of the disease (RRMS) at the beginning (4, 5). These patients are affected by attacks of unpredictable clinical relapses caused by inflammatory demyelinating lesions in the CNS, resulting in a complete or partial recovery of the neurological symptoms. After several years, the majority of these patients if untreated will develop secondary progressive MS (SPMS), where the neurological function decreases over time independent of relapse activity (6, 7). About 10-15% of the patients do not have relapses during the course of the disease. In these patients, the disease already begins with a gradual increase in neurological symptoms. This is called primary progressive progression (PPMS). Often a spastic gait disorder develops over the years, more rarely a progressive cerebellar syndrome (8). Beyond this raw classification of disease courses, each MS patient presents with a very individual course of his MS.

Longitudinal course. As described, MS is characterized by a chronic and/or episodic course. PwMS require long-term phenotyping, monitoring and most often treatment with disease-modifying therapies (DMTs) (9). In the early stages of MS, the damage occurring in the brain can still be compensated by the so-called neurological reserve. This compensatory mechanism explains why, on the one hand, early-stage MS is often not diagnosed promptly and, on the other hand, is often not taken seriously enough, especially with regard to negative long-term consequences (10–12). As the disease progresses, the neurological reserve decreases, especially if MS activity is not adequately treated (12). Since therapy started early in the course of the disease has an inhibitory effect on the progression of MS, it should be diagnosed and treated without any delays (13, 14).

Heterogeneous course and different dimensions. MS is popularly known as the "disease of a thousand faces" because MS lesions and other abnormalities can occur in the whole CNS usually leading to a variety of neurological deficits including fatigue, visual and bladder problems, pain, spasticity, reduced mobility and sexuality as well as psychological conditions such as depression (15–17). Due to this heterogeneity and the intra-individual unpredictable and inter-individually quite variable course, the diagnosis, phenotyping and monitoring of MS is very challenging (18, 19). The multidimensional disease characteristics of each patient should be made quantifiable to allow phenotyping of the individual disease characteristics and long-term monitoring of these parameters (20). This leads to a large amount of multidimensional data.

Multidimensional data. When quantifying MS, it is necessary to distinguish between different dimensions and perspectives. Starting from neurological-clinical parameters, they range from quantitative assessment of individual neurological functional systems (e.g. cognition, gait analysis), through imaging (MRI, ocular coherence tomography (OCT)), electrophysiological methods and the inclusion of patient-reported outcomes (PRO), up to new molecular and digital biomarkers (20). This data can be obtained in the setting of clinical trials or in real world practice, which represents also differences in its collection, volume, veracity and availability. To do justice to the complexity of MS, these parameters must be integrated into detailed individual patient charts as well as into large databases in order to be able to analyze them meaningfully.

Increasing number of potential therapeutic interventions. The number of treatment options that intervene in the immune system on different levels can modify disease is increasing (19, 21–25). This growing availability of DMTs is broadening the treatment options towards a more individualized therapy (24). Different mechanisms of action and intervention strategies are linked to individual treatments (26–29). On treatment, the monitoring of MS disease activity is key to achieve optimal outcomes in order to initiate a therapy change or escalation in time in case of an insufficient response (10, 30).

Therefore, the chronic, heterogenic and multifocal "disease of a thousand faces" requires a complex, ubiquitous and differentiated as well as adaptive diagnosis, monitoring and treatment strategy. This strategy should be personalized and tailored to the individual needs and disease course of the patient and be continuously adjusted (25).

Precision Medicine for People With Multiple Sclerosis

An emerging approach towards personalized treatment is precision medicine, or, as an older term, personalized medicine

(31–35), that takes into account individual variability in genes, environment, and lifestyle for each person (32, 36–42). Precision medicine covers diagnosis, treatment and management to achieve better patient outcomes (43). Through precision medicine, it is possible to break down the complexity of the disease. The patterns and inter-individual variability can be better understood. Thereby, precision medicine presents a framework for developing targeted treatment for individual patients by combining the demographic and clinical information, biomarkers and medical imaging data (44–47). Existing developments in precision medicine (44, 48–50) demonstrate that complex health-related big data of high quality are necessary, including lifestyle, nutrition, genetics, and environmental factors besides clinical, para-clinical, imaging and immunological or neurobiological parameters, which have to be analyzed and integrated in diagnosis, treatment and monitoring processes. To obtain big data and capture the bigger picture of a given individual on the way to precision medicine, Fagherazzi et al. recommend the method of "deep digital phenotyping", which is a combination of deep phenotyping by collecting biomedical data in the real world and digital phenotyping by collecting digital biomarkers (42, 44, 51–53).

In the patient´s perspective, a more transparent disease understanding can enable the patient to take a more active role in decision-making, following the concept of patient empowerment (54). Better understanding and involvement of patients in therapeutic decision making leads to better treatment adherence, which is associated with higher efficacy and lower healthcare costs (55). Ultimately, all patients would have the opportunity to query their own data interpreted in the context of the world's largest reference cohort and the latest data on available therapeutic options (56).

In relation to MS, deciding which therapy to use in a particular patient requires careful analysis of the patient's disease course for high-risk factors for early progression, consideration of the efficacy and safety profile for a potential therapy, and a patient's lifestyle and expectations (57). This is the only way to improve the precision of management for each patient, to improve prognosis and to establish an evidence-based framework for predicting response to treatment and personalized monitoring of patients. Precision medicine for pwMS involves the classification of disease subtypes based on underlying biology, not just clinical phenotype, and the development of predictive models that incorporate the integration of clinical, biological and molecular as well as current and emerging imaging markers with an understanding of the impact of the disease on the lives of individual patients (58–63). A complex data set could be the base of the DTMS as part of a digital data cloud that tries to simulate the same or very similar characteristics in terms of health status, risk factors and disease development as the real-world MS patient (43, 45).

WHAT ARE DIGITAL TWINS?

Origin and Concept Of Digital Twins

The concept of a "twin strategy" was generated from NASA's Apollo program, which build two real identical space vehicles. One was launched onto the air space, the other stayed on Earth to mirror the conditions of the launched one (64). The first mention of the term "digital twin" can be traced back to the year 2003 when Grieves mentioned it in the context of manufacturing (64–66). Initially, the space industry was primarily concerned with the topic of DT. In 2012, the NASA and the U.S. Air Force jointly published a paper about the DT, which stated the DT was the key technology for future vehicles. After that, the number of research studies on DT in aerospace has increased and the DT was introduced into more fields such as automotive, oil and gas as well as health care and medicine. Examples are online operation monitoring of process plants, traffic and logistics management, dynamic data assimilation enabled weather forecasting, real-time monitoring systems to detect leakages in oil and water pipelines, and remote control and maintenance of satellites or space-stations. For instance, Singapore is developing a digital copy of the entire city to monitor and improve utilities (67). In recent years, the DT has been described more and more as a promising technology and it is expected that DTs will develop very strongly in the coming years and will bring a revolution in several industry sectors with the desire for online monitoring, increasing flexibility and personalized services (64). The availability of cheap sensors and communication technologies and the phenomenal success of technologies such as machine learning (ML) and AI, new developments in computer hardware as well as cloud and edge computing will rapidly drive the development of the DT (66).

Grieves (65) originally defined the DT in three dimensions: a physical entity, a digital counterpart and a connection that ties the two parts together. In most definitions, the DT is considered as a virtual representation that interacts with the physical object throughout its lifecycle and provides intelligence for evaluation, optimization, prediction, etc. (68–72). For instance, in the industrial sector the DT is used as an in silico presentation of technical applications in order to optimize them through computer simulations (67, 73, 74). As these definitions focus on three dimensions (physical, virtual, connection of them), Tao et al. added the two further dimensions data and services. The newly proposed definition can fuse data from both the physical and virtual aspects using DT data for more comprehensive and accurate information capture (64). Kritzinger et al. divide DT into three subcategories, depending on the level of data integration (75). Rasheed et al. present an example of a state-of-the-art DT of an offshore oil platform. The DT is continuously updated with sensor data almost in real time. The sensor data can be supplemented with synthetic data from simulators that provide physical realism at high spatio-temporal resolution. The DT not only provides real-time information for more informed decision-making, but can also make predictions about how the plant will develop or behave in the future. In an ideal environment, a DT is indistinguishable from a physical object in both appearance and behavior, with the added benefit of being able to make predictions about the future. In fact, the DT also offers the possibility for people to physically interact with the object using an avatar (66).

Overall, it must be noted that the topic of DTs is of such variety and complicated that it is almost impossible to cover all aspects as it has been covered by several reviews (66, 76–88). Up to now, there are currently no common methods, standards or norms for the development of DTs. In order to exploit the potential of DTs, there are still many challenges to be taken (66, 67).

Digital Twins in Health Care

Focusing on the possibilities of DTs, medicine and healthcare are the areas that are likely to benefit most from the concept of DTs (66). There are several reasons for this. First, the number of intelligent portable devices and the organized storage of big data of individuals and cohorts is increasing. Second, human and thus medical thinking will eventually reach the natural limits of speed, complexity and performance. For HCPs, the massive and constant increase of knowledge in healthcare (e.g., differentiated diagnostics, more personalized therapies, interaction risks, active ingredients) is almost impossible to cope within daily work. HCPs are limited by everyday circumstances such as tiredness, time pressure and emotions. Especially in hospitals HCPs are under cost and time pressure and cannot always make decisions based solely on medical factors. And third, there is an increasing need for personalized and targeted treatment. As a result, various tools that enable precision medicine and simulation of therapies as well as prognosis of disease progression will inevitably find their way into the everyday life of HCPs, as is the case for already established different (clinical) decision support systems (CDSS) (89). The integration of technology and medicine is thus the main driver for intelligent and networked health. In this context, the statistical modeling of big data poses a particular challenge. Classical methods that examine associations between individual variables and a diagnosis or a course of disease reach their limits with the large number of statistical tests required and are also unable to uncover complex interactions between several variables and modalities in real time. Statistical significance, until now the primary measure of group-based, mechanistic research, also loses significance when, due to large samples, even the smallest effects exceed the significance threshold and thus the connection between significance and (clinical) relevance fades. ML is the key to creating direct clinical benefit. ML involves algorithms that can learn to solve a specific task autonomously based on data. These algorithms do not need to be explicitly programmed and can thus generate novel solutions to complex problems and tasks. Although classical statistical methods are capable of both correlation discovery and prediction, ML methods are better suited for identifying patterns, constructing features, and making predictions from large, complex, and heterogeneous data because they are usable and generalizable across a variety of data types and allow analysis and interpretation across complex variables. ML methods thus complement and extend existing statistical methods and can be used in highly innovative areas such as omics, radio-diagnostics, drug discovery, and personalized treatment. Of course, ML methods also have their limitations. The success of a ML project depends on the number of observations, the number of features, the selection and parameterization of the features as well as the quality of the underlying data and the chosen algorithm for the model (90, 91). ML also represents a component of AI research and development. AI is a computer system that is able to integrate relevant information and make a rational and logical decision that leads to the best possible outcome.

ML is an important component of a modern DT in healthcare (92), that can be defined as a “virtual mirror of ourselves that allows us to simulate our personal medical history and state of health using data-driven analytical algorithms and theory-driven physical knowledge” (93) as well as to exploit the synergies resulting from their combination. That is, a DT uses the induction approach (statistical models that learn from data) and the deduction approach (mechanistic models that integrate multiscale knowledge and data) to provide accurate predictions of pathways to maintain or restore health (45). A DT consists of numerous dynamic and multidimensional parameters. Dynamic data means that the data from which the digital image of the patient is created are both historically available data and continuously updating and accumulating data from that person’s life, e.g., data on the medical condition, data on the person’s living environment, data on how a drug is tolerated or a therapy is accepted. The multidimensionality of the data arises from the many different sources from which the data come, such as monitoring data, data from the patient’s social milieu, data from sensors, or clinical data. The dynamic and multidimensional nature of the data collected also distinguishes DT from other classical approaches such as clinical decision support systems (CDSS). A CDSS is used to make recommendations for appropriate tests and procedures from historical electronic health record (EHR) data using diagnosis of a condition and analysis of symptoms to help HCPs make informed decisions. The recommendation is the main component of a CDSS, which can be recorded in medical documents or coded in software as algorithms and rules (94, 95). However, the DT is not just a pure data collection approach for recommendations; it also correlates these data with each other and uses algorithms to incorporate the data meaningfully and purposefully into a simulation process with defined clinical (and economic) goals (95). The ability to simulate and model medical devices as well as pharmaceutical treatments on the computer enables faster and more cost-effective development than under real conditions (45, 48), without any risk for patients: “Making mistakes on computer models instead of people” (96).

The use of DTs in medical care is still in its infancy. So far, only in a few areas of medicine, DTs were applied, such as oncology (97–99), geriatrics (100, 101), cardiology (45, 102–106), epidemic outbreaks (107), genomic medicine (48, 108), internal medicine (109, 110), orthopedics (111) and vascular medicine (112, 113). For example, Corral-Acero et al. present early steps of a DT in the field of cardiovascular medicine by describing the synergies between mechanistic and statistical models, the pillars of the DT (45). Topol describes “high-performance medicine” with the help of AI for HCPs in different disciplines like radiology, pathology, dermatology, ophthalmology, cardiology and gastroenterology (114) and gives an overview over selected reports of machine- and deep-learning algorithms to predict clinical outcomes and related parameters. Laaki et al. developed the prototype of a DT for real-time remote control of remote operations over mobile networks (81). Bruynseels et al. show how DTs are based on in-silico representations of an individual that dynamically reflect molecular status, physiological state and lifestyle over time (46).

Concrete implementations of digital twins can already be found for organs such as the heart, for example, by the French software company Dassault Systèmes (115) or by Siemens Healthineers in Germany (116). Siemens Healthineers has used data collected in a huge database of more than 250 million annotated images, reports and operational data. The AI-based DT model was trained to weave data about a heart's electrical properties, physical characteristics and structure into a 3D image. The technology was tested on 100 digital heart twins from patients treated for heart failure in a six-year study. Preliminary results of the comparison between the actual outcome and the predictions the computer made after analyzing DT status seemed promising. French startup Sim&Cure developed a DT system that virtualizes a patient-based aneurysm and surrounding blood vessels. After a patient with aneurysm is prepared for surgery, a DT represented by a 3D model of the aneurysm and surrounding blood vessels is created by processing a 3D rotational angiography image. The personalized DT allows surgeons to perform simulations and helps them gain an accurate understanding of the interactive relationship between the implant and the aneurysm. In less than five minutes, numerous implants can be assessed to optimize the procedure. Preliminary studies have shown promising results (117).

CONCEPT OF DIGITAL TWINS IN THE MANAGEMENT OF MULTIPLE SCLEROSIS

Our vision is generating and implementing DTs in management of MS in order to improve diagnosis, treatment and management strategies as well as patient participation and compliance. DTs are a revolutionary tool for an improved characterization and prediction of disease course and for deep clinical phenotyping of pwMS (118). In this regard, big data analysis *via* ML supports visualization of the DTMS at different stages of MS and enables further therapeutic decisions. There are no elaborated DTs yet, but there are starting points and perspectives. For instance, Walsh et al. use an unsupervised ML model to learn the relationships between covariates commonly used to characterize subjects and their disease progression in clinical trials in MS (118). Recently, a research group from Sofia University in Bulgaria performed a first exercise of simulation of DTs. Petrova-Antonova et al. developed a web-based DT platform for MS diagnosis and rehabilitation that consists of two components: a transactional application that automates tests for MS diagnosis and rehabilitation, and an analytic application that provides data aggregation, enrichment, analysis, and visualization that can be used in any instance of the transactional application to generate new knowledge and support decision making. However, the analytical application is currently undeveloped and subject to further research (119).

We consider that, due to the complexity and long-term nature of MS, a particularly large and multidimensional amount of data must be collected and organized for the construction of DTMS. These data must be of high quality, i.e. they must be collected correctly and represent the patient as accurately as possible. In addition to quality, a high quantity and frequency of data collection must also be achieved in the long term. To create DTMS and keep them updated with follow-up data, parameters related to the patients physiological status data (structured clinical data, para-clinical and multi-omics data, and patient-reported data) and to procedures (diagnostic workup, treatment, monitoring as integrated into personalized clinical pathways) should be collected, analyzed, visualized and correlated (**Figure 1**). The evolving and self-updating DTMS can be used simultaneously with ML algorithms to make smarter predictions and decisions as a learning health system (LHS) (120).

Patients Physiological Status Data

Patients' physiological status data content of DTMS includes structured clinical and para-clinical data, some of them as digital data, as well as multi-omics and patient reported data.

Structured clinical data are key parameters of deep clinical phenotyping and prerequisite for the data content of DTMS (30, 121). Taking the patient's history is traditionally the first important step in the evaluation of pwMS, which focuses on relapses and/or disease progression in the different neurological functional systems. Contextual parameters including lifestyle factors, comorbidities (122), psychological factors, emotions and sociodemographic factors (123–125) must also be recorded, assessed through the medical record and the conversation between physician and patient. There are attempts to standardize and quantitate MS relevant neurological history, such as e.g. the MSProDiscuss tool in the assessment of secondary disease progression (126). Further clinical evaluation e.g. by neurological examination is indispensable in MS for the quantitative measurement of the extent of the disorder, which is in turn required to find out how the disease is evolving and the influence the different forms of treatment are having on it. In recent years, the Expanded Disease Disability Scale (EDSS) has been an essential, irreplaceable scale in MS which has been improved in the past years by different approaches (127–129). However, other additional clinical instruments have been introduced to quantitate the different multidimensional aspects of MS as fatigue, cognition or walking function (130, 131). The Multiple Sclerosis Functional Composite (MSFC) provides a functional assessment of different key functions (upper and lower extremities, cognition) that is used more and more frequently in MS and has been proven to be highly sensitive in the evaluation of very important clinical trials. These complex data could allow clinical phenotyping of MS in terms of disease activity (132) or symptom-specific phenotypes (133). Because DTs are data-driven approaches, it is not advisable to assume that the same monitoring procedures already used by the clinician in everyday practice are sufficient to establish a model for comprehensive digital representation of pwMS. Therefore, a combination of different clinical outcome measures is highly recommended (134). Initiatives to standardize the collection of clinical data are on the way (135).

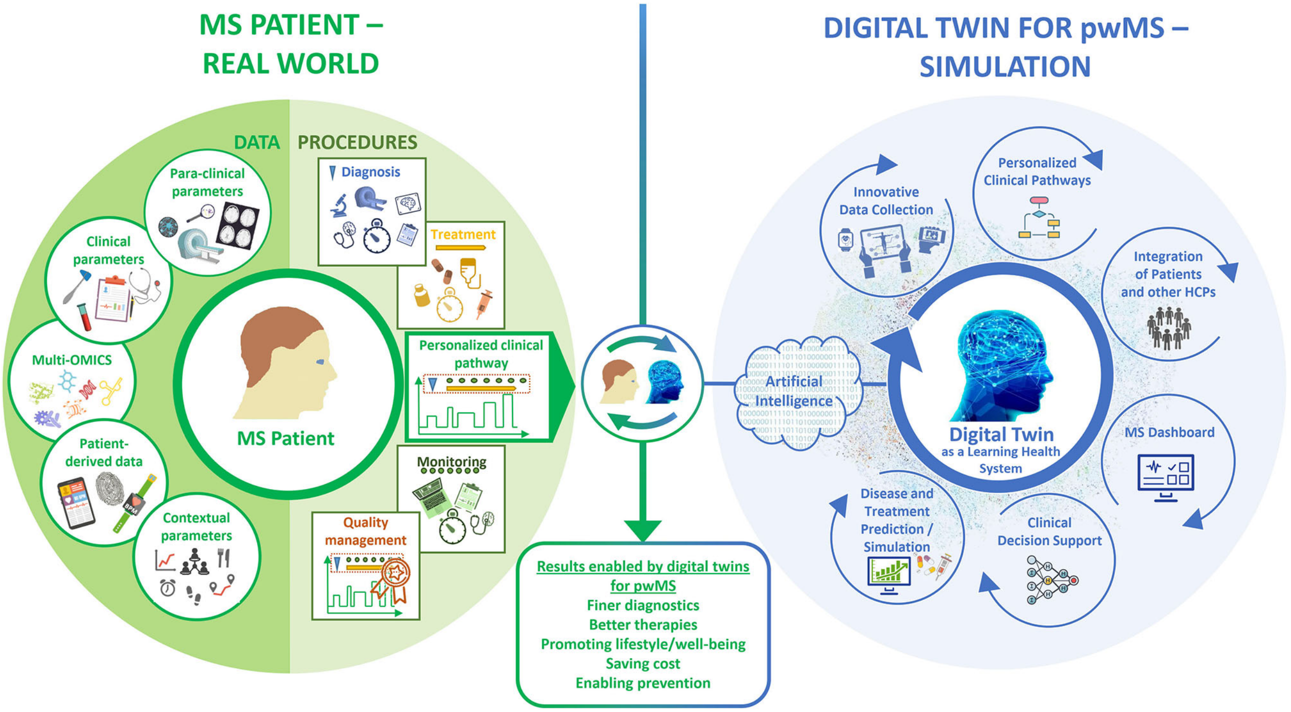


FIGURE 1 | Concept of a digital twin for pwMS.

Para-clinical data are of great importance for diagnosing, phenotyping and monitoring MS. Lab data ranging from standard laboratory to state of the art immunological or neurobiological parameters (136–139). Implementing standard lab data from clinical practice into a comprehensive approach of DTMS can complete the fundamental quest of real world evidence for individually improved treatment decisions and balanced therapeutic risk assessment (140, 141). As the MS disease process takes place in the CNS, analysis of cerebral spinal fluid is of high importance (123, 124). In addition to emerging immunological and neurobiological biomarkers, new technologies could be used for data collection for the DTMS as it has been described by Meyer zu Hörste (136) and will be described among multi-omics approaches. Neuronal destruction makers (e.g. neurofilament light chain) seem to be an excellent tool to measure subclinical MS disease activity in research and clinical studies (125, 142, 143), but final validation and transfer in clinical practice would be optimal in the setting of the multidimensional approach of DTMS.

The importance of CNS imaging has steadily increased in recent years and is expected to continue to grow in light of new sequencing techniques and applications related to pathophysiology and prediction (144, 145). As a biophysical technique for measuring magnetic properties and generating weighted images of relative tissue contrasts, MRI offers both volumetric and dynamic quantitative means of detecting pathological tissue changes. These represent a promising approach to optimizing MS management through *in vivo* monitoring in the assessment of the course of chronic diseases by recording their disease-related dynamics or treatment-induced effects (146, 147).

To implement imaging into DTMS, it is essential to standardize MRI acquisition (148, 149). The aim of this approach is to increase the sensitivity of MRI analysis to the smallest disease-related tissue changes. The acquisition of 3D-resolved sequences is important, as these, on the one hand, allow the free exploration of the image data by reformatting and, on the other hand, allow an optimal adaptation to the preliminary examination through modern 3D registration. In addition, only these 3D-resolved sequences form the basis for computer-assisted image data analysis and volumetric measurements, which should further increase precision in the future. Recent advances of CNS imaging could be probably transferred more easily into clinical practice by their integration into DTMS. Using this platform to put imaging data in context with other multidimensional data offers unique possibilities of validation and implementation. Thus, in future, quantitative MRI will enable a detailed characterization of brain tissue by generating a large number of numerical results (150). More than a thousand parameters can be generated if a detailed segmentation of the brain is considered, making group studies complex and inefficient by parametric techniques of data analysis (150). The large volume of MRI data can only be approached by AI, an essential tool of the DTMS (151). Finally, by measuring both volumetric and dynamic quantitative means (lesions and atrophy), different MRI phenotypes of individual patients can be described by MRI-categorization (152) which could be an important component of DTMS. In addition to MRI, data obtained through other imaging biomarkers such as OCT (153) or Positron emission tomography (154) can be used as well.

Digital phenotyping. Several clinical and para-clinical data can be collected digitally (digital phenotyping with digital biomarkers). Digital biomarkers are measures to collect objective data on biological (e.g., blood glucose, serum sodium), anatomical (e.g., mole size), or physiological (e.g., heart rate, blood pressure) parameter with the use of a biosensor (portable e.g. smartphones, wearable, and implantable devices), followed by the use of algorithms to transform these data into interpretable outcome measures (155–157). They are used for assessing e.g. cognitive function (158) or fatigue (159).

Sensor-based, portable measurement systems can be used both in the clinical setting and in the patients' individual everyday life (at home). In the clinical setting functional tests and gait analysis can be performed digitally. The Multiple Sclerosis Performance Test (MSPT) is a digital adaptation of the MSFC with additional elements added (160, 161) and measures health status *via* iPad with questionnaire on health status, processing speed with Processing Speed Test (PST) (162), manual skills with 9-Hole-Peg-Test (9-HPT) and walking speed with Timed 25-Foot-Walk (T25-FW) (160). Multidimensional gait analysis can be performed with measurement of walking speed (T25-FW), measurement of endurance [2-Minute Walk Test, 2MWT (163, 164)] and measurement of balance and gait quality on a sensor-based walking mat (GAITRite®-System, Mobility Lab-System) (131). For the digital measurement of data in patient-specific everyday life (at home) there are various patient apps such as Floodlight, diverse fitness tracker and health apps available (165, 166). They make it possible to collect realistic data relevant to everyday life *via* remote sensing in addition to the regular medical consultations. Thus, a more comprehensive insight into the patients' daily life as well as a more closely meshed progression monitoring is made possible. Clinical and para-clinical data (including lab and imaging data) are more and more collected in digital format and a standardized way which is an important step for integration in DTMS. A key role in the development of global standards of data related to patients or health cases is played by various organizations such as the Clinical Data Interchange Standards Consortium (CDISC), the Critical Path Institute (C-Path), and the Health Level 7 organizations (167). In clinical care, the development of digital neurological assessment tools such as Neurostatus-eEDSS and tablet-based MSPT, as well as real-time 3D motion capture systems for recording motor dysfunction in MS patients, play the most important role. The MS Data Alliance has already developed digital tools for aggregating, harmonizing, and sharing real-world data from multiple sources by creating a common data model. EHR also play a critical role in standardized and accurate digital documentation of clinical data, and several of these already exist, such as the MS BRIDGE, RC2NB, MSDS3D and MSBase EHR systems (168).

Multi-omics as innovative approach will have to be a part of the DTMS as well especially to increase knowledge about MS (169–171). The complex and dynamic processes in the neurobiological and immune networks are of significant importance in MS as in other chronic diseases. Advances in high-throughput "omics" technologies (e.g., genome, transcriptome, proteome, epigenome, metabolome) are enabling MS care to move from a "one-size-fits-all" toward a personalized approach analyzing the correlation of multi-omics with the clinical and para-clinical phenotypes of the individual MS patient (**Figure 2**). Multi-omics approaches involving large populations of pwMS and interrogating millions of markers with similar biochemical properties can help to elucidate the molecular mechanisms underlying MS and provide both potential biomarkers and pharmacological targets for a more detailed patient stratification and personalized treatments (172). Genomic and proteomic studies have sought to understand the molecular basis of MS and find biomarker candidates. Regarding genomic and proteomic studies, advances in next-generation sequencing and mass-spectrometry techniques have been of great importance to generate an unprecedented amount of relevant data (173). In order to study complex biological processes holistically, it is imperative to adopt an integrative approach. Multi-omics data should be combined to shed light on the interrelationships of the biomolecules involved and their functions. With the rinsing of high-throughput techniques and the availability of multi-omics data from a large number of samples, promising tools and methods for data integration and interpretation have been developed (174). In the field of MS, this strategy was successful for the development of novel data science techniques for exploring these large datasets to identify biologically relevant relationships and ultimately point towards useful biomarkers which have been discovered in recent years (124, 173).

Patient-reported data like questionnaire data complement the clinical data and complete the picture of the DTMS by including the patients' perspective of their disease. They are divided into patient reported outcomes (PRO) and patient reported experiences (PRE). PRO is an umbrella term for health outcomes that are directly and subjectively reported by patients without interpretation of the patients' response by a clinician or anyone else (175, 176). PRO are measured for outcomes like quality of life by the Quality of Life in Neurological Disorders (177, 178), and like walking and mobility skills by the Twelve Item MS Walking Scale (164, 179) or the Early Mobility Impairment Questionnaire (180). PRE measure "patient's perception of their personal experience of the healthcare they have received" (181). PRE measures assess patients' perception of their experience of the received healthcare collected through questionnaires (182). Efforts to standardize data are already underway. The PROMS (Patient Reported Outcomes for MS) initiative aims to identify PROs, including actively and passively delivered digital performance measures, to standardize outcomes in both research and clinical decision making (183).

Thus, model building for a DTMS already requires a comprehensive set of monitoring tools to be tested on a representative sample. To a certain extent, this also describes the scope of the instruments, which must later be applied to individual patients in practice in order to derive a comparable trajectories.

Procedures

An optimal management of pwMS requires the performance of certain procedures as e.g. assessments of clinical and para-

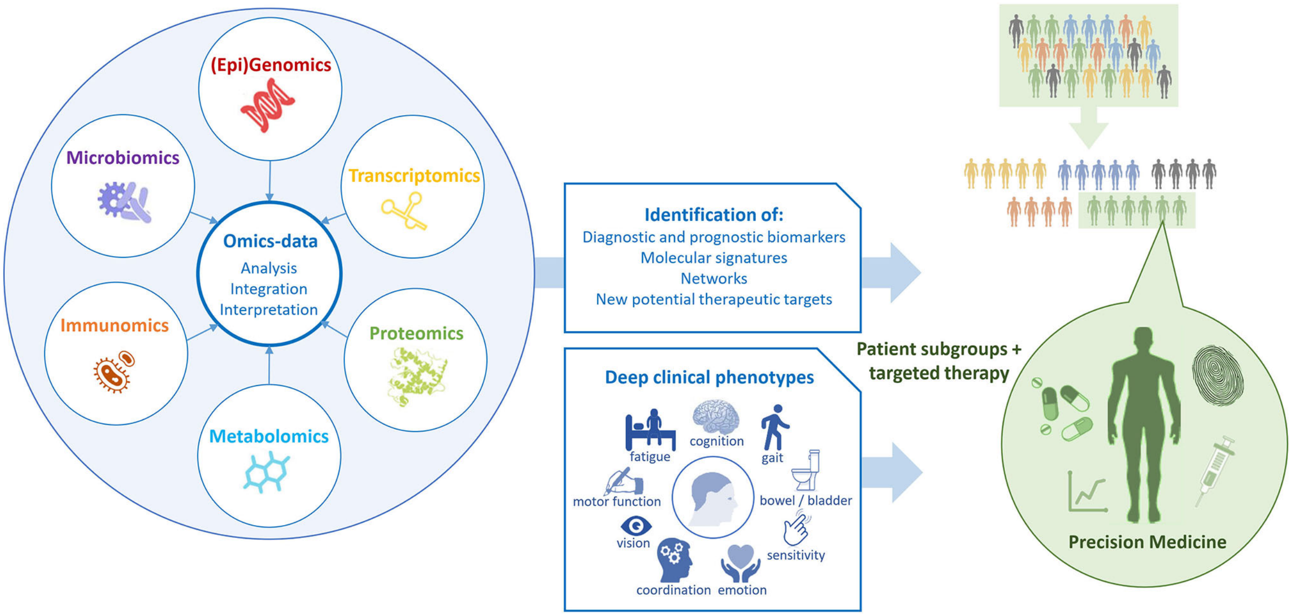


FIGURE 2 | Multi-omics for precision medicine.

clinical parameters at high quality and at defined time points. In addition to the more general and non-concrete guidelines related to standard clinical practice, the Brain Health Initiative has provided for the first time specific "core," "achievable," and "aspirational" time frames for individual treatment steps in diagnosis, treatment, and monitoring (14). Achieving these standards of MS management in the individual patient to increase quality of care for pwMS will be facilitated by integrating such procedural components of these clinical pathways into the DTMS.

Diagnosis of MS is based on defined diagnostic criteria [McDonald criteria (5)] and relies on various examination methods (184), none of which alone is capable of making the diagnosis of MS as the differential diagnosis is quite complex. The procedural component of diagnostic workup in DTMS will assist in collecting data in optimal time considering type and stage of disease, pertinent symptoms and comorbidities, time between the first referral to the neurologist and MRI, etc.

Treatment. The therapeutic management in MS includes DMTs, treatment of acute relapses, and symptomatic therapies, which are usually combined and individually adapted. In particular, the history and the stage of the disease, degree of disability, the primary symptomatology, form and dynamics of the course of the disease, age, gender and desire to have children, concomitant and previous diseases, concomitant and pre-medication as well as the individual life situation of the patient must be taken into account. The DTMS will assist in the selection and monitoring of individual treatments. In order to assess possible adverse events and reactions, individual treatments need a defined treatment-related clinical pathway including clinical and para-clinical assessments, which have to be integrated into the DTMS.

Monitoring. An optimal primary goal of MS therapy should be the achievement of no evidence of disease activity (NEDA) (9, 185). Specifically, this means the absence of relapses, new or enlarged lesions on MRI, clinical disability progression and loss of brain volume (=NEDA-4). The NEDA status has to be assessed by procedures of MS monitoring to detect disease progression and relapse as well as the monitoring of disease activity and symptoms. The importance of frequent high quality monitoring in routine clinical management of MS is pointed out by numerous authors with reference to various studies and the comprehensive data on the significance of relapses, early EDSS changes, and the role of MRI (186–188). As monitoring of MS is a lifelong challenge for patients and HCPs, its integration into DTMS will assist in keeping up this essential long term assessment.

Personalized clinical pathways that integrate these procedures are also included in the design of DTMS and should be available for the HCP and patient together to ensure the best possible outcome.

Construction of Digital Twins for Multiple Sclerosis

Prediction models based on statistical models already exist. For example, Stühler et al. and Kalincik et al. have investigated the individual response of pwMS to disease-modifying therapies using generalized linear models. However, in both studies, data density and quality were insufficient because, among other reasons, the cohorts were too small or there were data gaps in MRI data or data could not be comprehensively included (189–191). With the DTMS, all historically and currently available data should be continuously included in the analysis, if possible, to increase predictive power. In addition to the standardized and digitized parameters on patients' physiological status data and procedures, the available prior knowledge in the field of MS should also be included in the construction of the DTMS. In addition to existing guidelines (14), this also includes further expert knowledge from the practice of clinical care of pwMS as well as possible knowledge about factors that can positively or negatively influence the disease, e.g. comorbidities, nutrition, physical activity and cessation of smoking.

Before the DTMS is implemented in practice, it is essential to check which data are absolutely necessary to collect and how the data collection can be done in such a way that it burdens the patient and HCP as little as possible. This is also important from an economic point of view, as the collection of all the above-mentioned data types is associated with high costs. This examination could be done by different tools. Basically, a targeted literature review on parameters particularly frequently used for prognostic purposes would be necessary, which could be complemented by a survey among experts. Since the strength of ML methods lies in discovering hidden patterns, test runs of the DTMS with the integration of different parameters (classes) would be conceivable, the results of which would be tested in a representative sample. Some work already provides clues in this regard. As Pinto et al. have pointed out in their work on prediction of MS progression using ML methods, relevant clinical information may include EDSS, functional systems and CNS functions affected during relapses, as well as age and gender (192). In any case, data acquisition should be done digitally and in an automated manner, if at all possible, with a view to minimizing patient disruption. There is a need for further research in this area which data have been collected from patient and HCP.

USE CASES IN CARE OF MULTIPLE SCLEROSIS

DTMS perform a new kind of deep phenotyping by processing all data and procedural content in its complexity with innovative tools. Taking into account all previously defined medical and contextual parameters, which are very closely interwoven with the patient and his identity, the DTMS provides decision templates based on calculated probabilities. HCPs, patients, and all those involved in their care, have therefore an individualized roadmap of which examinations, tests, and therapies to pursue in the near future. In this process, the DTMS controls and monitors the entire disease management process and can correct any deviations. Thus, the DTMS is also a tool for measuring the process quality of a treatment. This results in a number of application scenarios that will fundamentally improve management of MS (**Figure 1**).

Innovative Data Collection

For linking large amounts of data from different sources, suitable interfaces and modular database systems should be available that can integrate different external systems. The ability of different systems to work together is called interoperability. To achieve interoperability and also flexibility, the use of an interoperability standard, such as HL7 FHIR (193), and standard interfaces, e.g. IHE XDS.b for Germany (194), should be ensured (195). This is where a MS portal such as the Integrated Care Portal Multiple Sclerosis (IBMS) (195) could be used, to which both patients and HCPs can contribute different types of data. Patient data collected *via* apps or questionnaires flow into the patient portal, which is part of a management system for MS. The HCP, in turn, can see this data in the system and enter content related to the data and processes there. In the further course, data enter the database continuously, which can be used for the DT.

Clinical Pathways

Clinical pathways are particularly suitable for the seamless care of chronically ill patients across various health sectors. They describe the entire path of patients during care (the "patient journey") and unite the multidisciplinary setting, the local conditions and the current state of evidence research (195). Clinical pathways define goals and milestones of care and support shared decision making between HCPs and patients by also providing patients with a picture of their stage of disease (30, 195–198).

As intelligent systems, DTs traverse the clinical pathway, serving as a guide for HCPs and patients through treatment with an individual roadmap. Integrated into clinical management systems, clinical pathways can thus also serve as quality assurance tools for HCPs and patients. In this way, patients can actively participate in the quality improvement of their treatment process. HCPs, in turn, have the opportunity to optimize treatment steps based on specific quality indicators. These quality indicators are derived from existing MS guidelines and consensus standards [e.g., the International Brain Health Initiative consensus standards (14)]. On the one hand, they address temporal concerns for diagnosis, treatment, and monitoring phases, e.g., the maximum time between initial presentation and the acquisition of an MRI. On the other hand, quality instruments are integrated to measure the assurance of desired outcomes for pwMS, e.g., whether patients who have mobility or fatigue issues are offered support (199) or whether patients experience coordinated care with clear and accurate information exchange (200). Defining and measuring quality indicators is the goal of the currently running project "Path-based Quality Management in MS Care" (QPATH4MS) at the MS Center Dresden (Germany).

MS Dashboard for Visualization

Visualization helps to present complex data in an understandable and clear way. The so-called MS dashboard visualizes high-dimensional disease characteristics and individual clinical pathways. The HCP can present the possibilities played out by means of the DT to the patient to discuss therapy options and clinical pathways with the patient. Through an adaptive display, it is possible to present the individual patient pathway, therapy options, treatment alternatives and the associated risks and challenges in a simplified form for the patient as a layperson and for the HCP as an expert. Within this framework, HCPs and patients can determine the ideal therapy and management of MS through shared decision-making. Thanks to the visualized simulation of the DT, the HCP has time to address all patients' questions and concerns in detail. Examples of existing dashboards for displaying individual patient data at a glance include the walking assessment dashboard as part of the multidimensional digital patient management system MSDS^{3D} (201, 202), showing the results of clinical multidimensional walking assessment and daily smart monitoring longitudinally (131), and the MS BioScreen, that integrates multiple dimensions of disease information: clinical evolution, therapeutic interventions, brain, eye, and spinal cord imaging, environmental exposures, genomics, and biomarker data (56, 203).

Integration of Patients and Other Healthcare Professionals

The visualization of the complex data involved in the medical and therapeutic decisions may foster the communication between HCPs and patients. This would support the involvement of patients in healthcare decisions and management of their disease. In this way, DTs also serve as a shared decision-making tool for HCPs and patients, who will play a much more active role in their own healthcare management in the future. For example, this could empower the patient to become an active member of the MS management team, from providing data (including data from biosensors, for example) to recording/tracking notable events and daily care to prognostic tools. As a result, a much more granular, continuous perspective on MS and its progression is provided, which would be more complete than traditional (brief and irregular) clinical assessments.

Clinical Decision Support System

A DT also acts as clinical decision support system (CDSS) that supports HCPs in clinical decision making by providing evidence-based medical knowledge and patient-related information (204, 205). The goal is to enable the HCP to make the best possible clinical decision for the patient, with the best possible chance of a positive outcome. CDSSs are often supported by ML-based algorithms. The ambiguous patterns of MS (e.g., in etiology, progression, clinical presentation, and response to drug therapies) make ML algorithms optimal tools to automate the detection of patterns and regularities in MS data. CDSSs are very beneficial in the context of MS, but are not yet well established. There is an increasing need for CDSSs in MS to help HCPs make the right decision among multiple alternatives in time (206).

Simulation and Prediction of Disease and Treatment Outcomes

Modeling the course of MS, especially predicting progression, is challenging due to the complexity of the outcomes and its varying course. The DT offers the possibility of predicting several probable disease courses and provides models for estimating possible treatment effects for individual patients. Taking into consideration all of a patient's individual parameters, potential side effects, costs incurred, and individual circumstances and patient satisfaction, the DT can suggest the option with the highest benefit for the patient. There are initial approaches to predicting disease progression using ML. For instance, Pinto et al. used clinical information to develop a ML system to explore the disease evolution in pwMS in terms of conversion from RRMS to SPMS. EDSS score, majority of functional systems, affected functions during relapses, and age at onset were described as the most predictive features (192). Zhao et al. found that support vector machines incorporating short-term clinical and brain MRI data were better at predicting disease progression of MS and selecting patients for more aggressive treatments than logistic regression methods (207). Later, Zhao et al. compared common ML algorithms and so-called ensemble learning approaches. The latter were more effective and robust compared with single algorithms and offered increased accuracy for predicting disease progression of MS. Of the variables evaluated, EDSS, pyramidal function, and ambulation index were the most common predictors in predicting MS disease progression (208). Another study suggested that the concentration of serum cytokines could be used as prognostic marker for the prediction of MS (209). Data-driven subtyping and staging of MS could better predict subsequent clinical course and response to treatment compared with clinical classification or baseline EDSS. Data-driven subtyping has the potential to prospectively improve patient outcomes.

DTs help to understand disease's dynamics and thus, advise HCPs on medication intake. With regard to drugs, it is quite conceivable that in the future clinical trials will also be conducted only with the help of DTs and no longer with the patients themselves.

From all that is known so far, the DT is a Learning Health System (LHS). LHS fuse healthcare delivery with research, data science, and quality improvement processes. The LHS cycle begins and ends with HCP-patient interactions and strives for continuous improvement in healthcare quality, outcomes, and efficiency (210). Based on the constantly new data collected through continuous monitoring and provided by the patient from the real world, the DT generates new knowledge, which in turn flows into the patient's further treatment, which is thus continuously improved. The parameter data continuously flow into the calculations of the DT – with each piece of information, the phenotype can be described more precisely. The therapy can thus be continuously adapted to the patient's disease state and life circumstances.

CHALLENGES OF DIGITAL TWINS IN HEALTH CARE

The use of DTMS promises to improve clinical decision making for individual patients, enhance patient communication, and improve quality of care. However, no uniform methods, standards, or norms yet exist for the development of DTs, and many challenges remain to unleash the potential of DTs (49, 66).

Data Quality, Data Management and Algorithm Design

Poor or missing data and information can lead to improper models and incorrect recommendations (trash in, trash out). In order for the DT to be statistically indistinguishable from its real-world counterpart, the data on which the DT is based must be of high quality and represent the patient as completely as possible (118, 120). Data quality in the broader sense also includes the standardized collection or standardization of data to ensure their reliability and to enable longitudinal and cross-sectional comparisons of data. In this context, data should preferably be carried out in digital form or at least recorded digitally instead of in paper form in order to facilitate standardization and thus comparability. There is currently no generally accepted, standardized scheme for the collection, documentation, and evaluation of data in MS, although recommendations and guidelines from various expert groups exist, which have already been described in the sections on patients physiological status data and procedures (135). For the purpose of generating a sufficiently large amount of data describing pwMS in a standardized multidimensional manner, many years of multicenter data acquisition are required. Only on this basis is it possible to create the necessary "critical mass" of data in the required density to enable long-term estimation of therapeutic outcomes. In addition, multidimensional and unstructured large data sets must first be structured and then integrated into meaningful algorithms before meaningful models can be created (45, 95, 114). It should also be noted that the results of ML algorithms are usually based on a large number of parameters and criteria that can no longer be reproduced or fully understood by humans (135). Even if the models produce solid predictions, it may be impossible to deduce why they make good prediction.

Data Privacy and Data Security

Before DTs are created, it is essential to clarify who owns which data at what point in time and for how long, who has access to it under which conditions and for how long, who actually owns the "end product" of the DT, and who can use it and under which conditions. It is imperative that suitable governance structures be created for this purpose. Furthermore, data security is very important to avoid data gaps that could potentially be used for hacker attacks to the detriment of patients. It is also necessary to ensure the protection of privacy, which becomes more and more difficult with the increasing functionality of techniques. Patients must also be confident that their data is secure, transparent and accessible to them. Otherwise, the collection of patient data could increase mistrust rather than confidence in health systems. Simply providing technological advances is not enough, it is also necessary to ensure that it serves to improve well-being. Therefore, data privacy and transparency of data use must be respected with the full consent of patients. Informed consent should explicitly state the purposes for which the data collected from patients will be used (49, 93, 120, 211).

Ethical Concerns

DT models could exaggerate racial and other bias (46, 212) and could lead to or reinforce inequalities in health care (46): if a group is misrepresented in the data used to create models, this group may receive suboptimal treatment (213). An example shows that a computer model classifies patients with a history of asthma who have pneumonia as patients with a lower risk of mortality than those who have pneumonia only. However, the context was completely ignored, namely that this is an artifact of clinicians admitting and treating such asthma patients earlier and more aggressively (212). Another important ethical issue related to predicting the course of disease is whether and in what way the prognosis should be communicated to patients. How does a patient deal with the knowledge that, according to the prognosis, he or she will soon be in a wheelchair, for example? Do patients have a right to "not know"? In addition, the extent to which patients will be able to decide autonomously what is good or bad for them, and to what extent this will be determined by the algorithms that claim to propose the most optimal solution based on the available data, needs to be reconsidered. In this context, "dataism" could become a new form of medical paternalism. Patients must therefore develop an appropriate relationship with their personal DT and develop the ability to make informed decisions in the face of strong data-driven personalized models (46).

Individual Concerns and Trust in Applications of AI

The role of humans or users of AI applications should not be underestimated, and trust is a crucial factor in this context (120, 214). The fear of new not-yet-established technologies like AI is a barrier to trust (120). HCPs may not trust the decisions of machines if they do not understand the involved algorithms. Additionally, HCPs could experience fear of being replaced by machines. However, AI will not replace the HCP (114), but will support and provide more time for consultation with the patient – one of the crucial aspects of medical care (215). Decisions based on AI can help the HCP make good decisions, if they "keep human intelligence up to date and take into account the social, clinical and personal context" (212). In the case that the DT's recommendations contradict his or her own, the HCP must dispose of an action plan for further decision-taking. Otherwise, more data can contribute to the uncertainty of the medical thinking.

In order to establish the concept of the DT despite all the challenges mentioned, guidelines, gold standards, benchmark tests and governmental legislation, as has been achieved in Estonia, are therefore necessary (45, 114). Before using DTs in patient care, it is imperative that targeted studies, publication of results in peer-reviewed journals and clinical validation in a real-world environment are carried out (114). Nevertheless, HCP should proactively guide, supervise and monitor the introduction of DTs as partners in patient care (212).

DISCUSSION

With the development of a DTMS, it is possible to improve clinical decision-making for individual patients, patient communication, shared decision-making, and thus quality of care. Before DTs can be used in patient care, they must be validated by studies and experts, as well as by real-world investigations to show the effectiveness and safety of their methods. In addition, there are still a number of challenges to overcome on the road to using DTs, such as ensuring data security and privacy and the accuracy of the data on which the

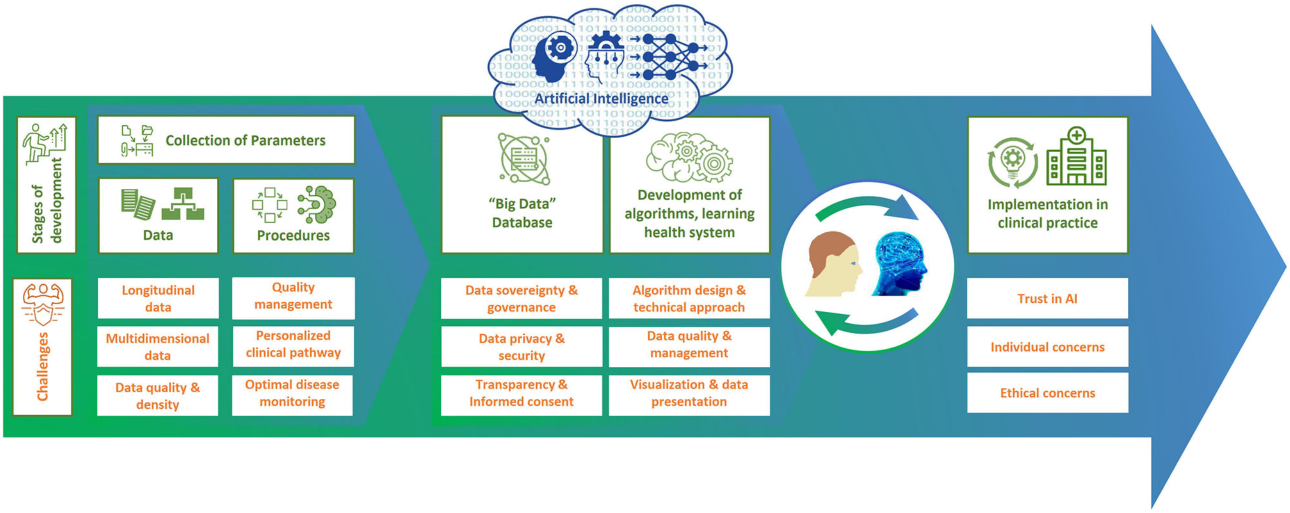

Artificial Intelligence
Stages of development
Collection of Parameters
Data
Procedures
"Big Data" Database
Development of algorithms, learning health system
Implementation in clinical practice
Challenges
Longitudinal data
Quality management
Multidimensional data
Personalized clinical pathway
Data quality & density
Optimal disease monitoring
Data sovereignty & governance
Data privacy & security
Transparency & Informed consent
Algorithm design & technical approach
Data quality & management
Visualization & data presentation
Trust in AI
Individual concerns
Ethical concerns

FIGURE 3 | Time flow for digital twins.

DT is based (**Figure 3**). It should also not be underestimated that the development of a DTMS is very complex and therefore expensive and may also increase the complexity of monitoring in clinical practice. Therefore, further research should be included in the development to inform which data contribute most to predictability, how this predictability can be assessed, and how this approach can be feasibly and cost-effectively integrated into health care. Further work will also be required to see whether and how predictive models can be constructed. However, a basic DTMS can serve as a starting point that will grow and evolve over time. During this process, the HCP should proactively guide, oversee, and monitor the introduction of DTMS as partners in patient care. By analyzing all possible factors of MS, DTMS will help make precision medicine and patient-centered care a reality in everyday life. This will ultimately refine diagnostics and monitoring, improve therapies and patient well-being, save economic costs, enable prevention, expand treatment options and empower patients.

AUTHOR CONTRIBUTIONS

IV and TZ designed and conceptualized paper as well as revised the manuscript for intellectual content. HI, AD, RH and AD revised the manuscript for intellectual content. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We are grateful that the IBMS and QPATH4MS projects were funded by the ERDF.

REFERENCES

1. Filippi M, Rocca MA, Ciccarelli O, De Stefano N, Evangelou N, Kappos L, et al. MRI Criteria for the Diagnosis of Multiple Sclerosis: MAGNIMS Consensus Guidelines. *Lancet Neurol* (2016) 15(3):292–303. doi: 10.1016/S1474-4422(15)00393-2
2. Goldenberg MM. Multiple Sclerosis Review. *P T Peer-Rev J Formulary Manage* (2012) 37(3):175–84.
3. Huang WJ, Chen WW, Zhang X. Multiple Sclerosis: Pathology, Diagnosis and Treatments (Review). *Exp Ther Med* (2017) 13(6):3163–6. doi: 10.3892/etm.2017.4410
4. Khurana V, Sharma H, Medin J. Estimated Prevalence of Secondary Progressive Multiple Sclerosis in the USA and Europe: Results From a Systematic Literature Search (P2.380). *Neurology* (2018) 90(15 Supplement): P2.380.
5. Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, et al. Diagnosis of Multiple Sclerosis: 2017 Revisions of the McDonald Criteria. *Lancet Neurol* (2018) 17(2):162–73. doi: 10.1016/S1474-4422(17)30470-2
6. Weinshenker BG. The Natural History of Multiple Sclerosis: Update 1998. *Semin Neurol* (1998) 18(3):301–7. doi: 10.1055/s-2008-1040881
7. Lublin FD, Baier M, Cutter G. Effect of Relapses on Development of Residual Deficit in Multiple Sclerosis. *Neurology* (2003) 61(11):1528–32. doi: 10.1212/01.WNL.0000096175.39831.21
8. Thompson AJ, Montalban X, Barkhof F, Brochet B, Filippi M, Miller DH, et al. Diagnostic Criteria for Primary Progressive Multiple Sclerosis: A Position Paper. *Ann Neurol* (2000) 47(6):831–5. doi: 10.1002/1531-8249(200006)47:6<831::AID-ANA21>3.0.CO;2-H
9. Weinstock-Guttman B, Medin J, Khan N, Korn JR, Lathi E, Silversteen J, et al. Assessing 'No Evidence of Disease Activity' Status in Patients With Relapsing-Remitting Multiple Sclerosis Receiving Fingolimod in Routine Clinical Practice: A Retrospective Analysis of the Multiple Sclerosis Clinical and Magnetic Resonance Imaging Outcomes in the USA (Ms-Mrius) Study. *CNS Drugs* (2018) 32(1):75–84. doi: 10.1007/s40263-017-0482-4
10. Ziemssen T, Derfuss T, de Stefano N, Giovannoni G, Palavra F, Tomic D, et al. Optimizing Treatment Success in Multiple Sclerosis. *J Neurol* (2016) 263(6):1053–65. doi: 10.1007/s00415-015-7986-y
11. Inojosa H, Proschmann U, Akgun K, Ziemssen T. A Focus on Secondary Progressive Multiple Sclerosis (SPMS): Challenges in Diagnosis and Definition. *J Neurol* (2021) 268(4):1210–21. doi: 10.1007/s00415-019-09489-5

12. Inojosa H, Proschmann U, Akgün K, Ziemssen T. Should We Use Clinical Tools to Identify Disease Progression? *Front Neurol* (2021) 11:628542–. doi: 10.3389/fneur.2020.628542

13. Montalban X, Gold R, Thompson AJ, Otero-Romero S, Amato MP, Chandraratna D, et al. Ectrims/Ean Guideline on the Pharmacological Treatment of People With Multiple Sclerosis. *Mult Scler* (2018) 24(2):96–120. doi: 10.1177/1352458517751049

14. Hobart J, Bowen A, Pepper G, Crofts H, Eberhard L, Berger T, et al. International Consensus on Quality Standards for Brain Health-Focused Care in Multiple Sclerosis. *Mult Scler* (2018) 25(13):1809–18. doi: 10.1177/1352458518809326

15. Reich DS, Lucchinetti CF, Calabresi PA. Multiple Sclerosis. *New Engl J Med* (2018) 378:169–80. doi: 10.1056/NEJMra1401483

16. Yaldizli Ö, Kappos L. *Klinische Grundlagen Der Multiplen Sklerose*. S Egli, editor. Berlin and New York: Springer (2011).

17. Kip M, Schönfelder T, Bleß H. Versorgungssituation in Deutschland. In: *Weißbuch Multiple Sklerose* Berlin: Springer (2016).

18. Alkhawajah M, Oger J. When to Initiate Disease-Modifying Drugs for Relapsing Remitting Multiple Sclerosis in Adults? *Multiple Sclerosis Int* (2011) 2011:724871. doi: 10.1155/2011/724871

19. Merkel B, Butzkueven H, Traboulsee AL, Havrdova E, Kalincik T. Timing of High-Efficacy Therapy in Relapsing-Remitting Multiple Sclerosis: A Systematic Review. *Autoimmun Rev* (2017) 16(6):658–65. doi: 10.1016/j.autrev.2017.04.010

20. Ziemssen T. Dem MS-Phänotyp Auf Der Spur. *DNP Der Neurologe Psychiater* (2019) 20(5):33–6. doi: 10.1007/s15202-019-2277-6

21. Ziemssen T, Medin J, Couto CA, Mitchell CR. Multiple Sclerosis in the Real World: A Systematic Review of Fingolimod as a Case Study. *Autoimmun Rev* (2017) 16(4):355–76. doi: 10.1016/j.autrev.2017.02.007

22. Tacchella A, Romano S, Ferraldeschi M, Salvetti M, Zaccaria A, Crisanti A, et al. Collaboration Between a Human Group and Artificial Intelligence can Improve Prediction of Multiple Sclerosis Course: A Proof-of-Principle Study. *F1000Research* (2017) 6:2172. doi: 10.12688/f1000research.13114.1

23. Marziniak M, Ghorab K, Kozubski W, Pfleger C, Sousa L, Vernon K, et al. Variations in Multiple Sclerosis Practice Within Europe - Is it Time for a New Treatment Guideline? *Mult Scler Relat Disord* (2016) 8:35–44. doi: 10.1016/j.msard.2016.04.004

24. Ohlmeier C, Gothe H, Haas J, Osowski U, Weinhold C, Blauwitz S, et al. Epidemiology, Characteristics and Treatment of Patients With Relapsing Remitting Multiple Sclerosis and Incidence of High Disease Activity: Real World Evidence Based on German Claims Data. *PloS One* (2020) 15(5): e0231846. doi: 10.1371/journal.pone.0231846

25. Ziemssen T, Thomas K. Treatment Optimization in Multiple Sclerosis: How do We Apply Emerging Evidence? *Expert Rev Clin Immunol* (2017) 13 (6):509–11. doi: 10.1080/1744666X.2017.1292135

26. Sellner J, Rommer PS. Immunological Consequences of "Immune Reconstitution Therapy" in Multiple Sclerosis: A Systematic Review. *Autoimmun Rev* (2020) 19(4):102492. doi: 10.1016/j.autrev.2020.102492

27. Sellner J, Rommer PS. A Review of the Evidence for a Natalizumab Exit Strategy for Patients With Multiple Sclerosis. *Autoimmun Rev* (2019) 18 (3):255–61. doi: 10.1016/j.autrev.2018.09.012

28. D'Amico E, Zanghì A, Gastaldi M, Patti F, Zappia M, Franciotta D. Placing CD20-Targeted B Cell Depletion in Multiple Sclerosis Therapeutic Scenario: Present and Future Perspectives. *Autoimmun Rev* (2019) 18(7):665–72. doi: 10.1016/j.autrev.2019.05.003

29. Montes Diaz G, Hupperts R, Fraussen J, Somers V. Dimethyl Fumarate Treatment in Multiple Sclerosis: Recent Advances in Clinical and Immunological Studies. *Autoimmun Rev* (2018) 17(12):1240–50. doi: 10.1016/j.autrev.2018.07.001

30. Ziemssen T, Kern R, Thomas K. Multiple Sclerosis: Clinical Profiling and Data Collection as Prerequisite for Personalized Medicine Approach. *BMC Neurol* (2016) 16:124. doi: 10.1186/s12883-016-0639-7

31. Abrahams E. Right Drug-Right Patient-Right Time: Personalized Medicine Coalition. *Clin Trans Sci* (2008) 1(1):11–2. doi: 10.1111/j.1752-8062.2008.00003.x

32. Jameson JL, Longo DL. Precision Medicine–Personalized, Problematic, and Promising. *N Engl J Med* (2015) 372(23):2229–34. doi: 10.1056/NEJMsb1503104

33. National Research Council Committee on AFfDaNToD. The National Academies Collection: Reports Funded by National Institutes of Health. In: *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. Washington (DC: National Academies Press (US) Copyright © 2011, National Academy of Sciences (2011).

34. Sugeir S, Naylor S. Critical Care and Personalized or Precision Medicine: Who Needs Whom? *J Crit Care* (2018) 43:401–5. doi: 10.1016/j.jcrc.2017.11.026

35. Abrahams E, Ginsburg GS, Silver M. The Personalized Medicine Coalition. *Am J Pharmacogenom* (2005) 5(6):345–55. doi: 10.2165/00129785-200505060-00002

36. Toward Precision Medicine. *Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. NR Council, editor. Washington, DC: The National Academic Press (2011). p. 142.

37. Ashley EA. The Precision Medicine Initiative: A New National Effort. *JAMA* (2015) 313(21):2119–20. doi: 10.1001/jama.2015.3595

38. Collins FS, Varmus H. A New Initiative on Precision Medicine. *New Engl J Med* (2015) 372(9):793–5. doi: 10.1056/NEJMp1500523

39. Collins H, Calvo S, Greenberg K, Forman Neall L, Morrison S. Information Needs in the Precision Medicine Era: How Genetics Home Reference can Help. *Interactive J Med Res* (2016) 5(2):e13–e. doi: 10.2196/ijmr.5199

40. Conrad K, Shoenfeld Y, Fritzler MJ. Precision Health: A Pragmatic Approach to Understanding and Addressing Key Factors in Autoimmune Diseases. *Autoimmun Rev* (2020) 19(5):102508. doi: 10.1016/j.autrev.2020.102508

41. Jain KK. *Textbook of Personalized Medicine*. Switzerland: Springer, Cham (2021).

42. König IR, Fuchs O, Hansen G, von Mutius E, Kopp MV. What is Precision Medicine? *Eur Respir J* (2017) 50(4):1700391. doi: 10.1183/13993003.00391-2017

43. Hansen MR, Okuda DT. Precision Medicine for Multiple Sclerosis Promotes Preventative Medicine. *Ann New Y Acad Sci* (2018) 1420(1):62–71. doi: 10.1111/nyas.13846

44. Fagherazzi G. Deep Digital Phenotyping and Digital Twins for Precision Health: Time to Dig Deeper. *J Med Internet Res* (2020) 22(3):e16770. doi: 10.2196/16770

45. Corral-Acero J, Margara F, Marciniak M, Rodero C, Loncaric F, Feng Y, et al. The 'Digital Twin' to Enable the Vision of Precision Cardiology. *Eur Heart J* (2020) 41:4556–64. doi: 10.1093/eurheartj/ehaa159

46. Bruynseels K, Santoni de Sio F, van den Hoven J. Digital Twins in Health Care: Ethical Implications of an Emerging Engineering Paradigm. *Front Genet* (2018) 9:31–. doi: 10.3389/fgene.2018.00031

47. Mahler M, Martinez-Prat L, Sparks JA, Deane KD. Precision Medicine in the Care of Rheumatoid Arthritis: Focus on Prediction and Prevention of Future Clinically-Apparent Disease. *Autoimmun Rev* (2020) 19(5):102506. doi: 10.1016/j.autrev.2020.102506

48. Björnsson B, Borrebaeck C, Elander N, Gasslander T, Gawel DR, Gustafsson M, et al. Digital Twins to Personalize Medicine. *Genome Med* (2019) 12 (1):4–. doi: 10.1186/s13073-019-0701-3

49. Ienca M, Ferretti A, Hurst S, Puhan M, Lovis C, Vayena E. Considerations for Ethics Review of Big Data Health Research: A Scoping Review. *PloS One* (2018) 13(10):e0204937. doi: 10.1371/journal.pone.0204937

50. Ziegelstein RC. Personomics: The Missing Link in the Evolution From Precision Medicine to Personalized Medicine. *J Personalized Med* (2017) 7 (4):11. doi: 10.3390/jpm7040011

51. Robinson PN. Deep Phenotyping for Precision Medicine. *Hum Mutat* (2012) 33(5):777–80. doi: 10.1002/humu.22080

52. Dorsey ER, Omberg L, Waddell E, Adams JL, Adams R, Ali MR, et al. Deep Phenotyping of Parkinson's Disease. *J Parkinson's Dis* (2020) 10(3):855–73. doi: 10.3233/jpd-202006

53. Delude CM. Deep Phenotyping: The Details of Disease. *Nature* (2015) 527 (7576):S14–5. doi: 10.1038/527S14a

54. Rieckmann P, Centonze D, Elovaara I, Giovannoni G, Havrdova E, Kesselring J, et al. Unmet Needs, Burden of Treatment, and Patient Engagement in Multiple Sclerosis: A Combined Perspective From the MS in the 21st Century Steering Group. *Mult Scler Relat Disord* (2018) 19:153–60. doi: 10.1016/j.msard.2017.11.013

55. Lin X, Yu M, Jelinek GA, Simpson-Yap S, Neate S, Nag N. Greater Engagement With Health Information Is Associated With Adoption and Maintenance of Healthy Lifestyle Behaviours in People With MS. *Int J Environ Res Public Health* (2020) 17(16):5935. doi: 10.3390/ijerph17165935
56. Gourraud PA, Henry RG, Cree BA, Crane JC, Lizee A, Olson MP, et al. Precision Medicine in Chronic Disease Management: The Multiple Sclerosis Bioscreen. *Ann Neurol* (2014) 76(5):633–42. doi: 10.1002/ana.24282
57. Brück W, Gold R, Lund BT, Oreja-Guevara C, Prat A, Spencer CM, et al. Therapeutic Decisions in Multiple Sclerosis: Moving Beyond Efficacy. *JAMA Neurol* (2013) 70(10):1315–24. doi: 10.1001/jamaneurol.2013.3510
58. Gafson A, Craner MJ, Matthews PM. Personalised Medicine for Multiple Sclerosis Care. *Mult Scler* (2017) 23(3):362–9. doi: 10.1177/1352458516672017
59. Chitnis T, Prat A. A Roadmap to Precision Medicine for Multiple Sclerosis. *Multiple Sclerosis J* (2020) 26(5):522–32. doi: 10.1177/1352458519881558
60. Bose G, Freedman MS. Precision Medicine in the Multiple Sclerosis Clinic: Selecting the Right Patient for the Right Treatment. *Mult Scler* (2020) 26 (5):540–7. doi: 10.1177/1352458519887324
61. Comabella M, Sastre-Garriga J, Montalban X. Precision Medicine in Multiple Sclerosis: Biomarkers for Diagnosis, Prognosis, and Treatment Response. *Curr Opin Neurol* (2016) 29(3):254–62. doi: 10.1097/WCO.0000000000000336
62. Golan D, Staun-Ram E, Miller A. Shifting Paradigms in Multiple Sclerosis: From Disease-Specific, Through Population-Specific Toward Patient-Specific. *Curr Opin Neurol* (2016) 29(3):354–61. doi: 10.1097/WCO.0000000000000324
63. Pulido-Valdeolivas I, Zubizarreta I, Martinez-Lapiscina EH, Villoslada P. Precision Medicine for Multiple Sclerosis: An Update of the Available Biomarkers and Their Use in Therapeutic Decision Making. *Expert Rev Precis Med Drug Dev* (2017) 2(6):345–61. doi: 10.1080/23808993.2017.1393315
64. Tao F, Zhang M, Nee AYC. Chapter 1 - Background and Concept of Digital Twin. In: F Tao, M Zhang and AYC Nee, editors. *Digital Twin Driven Smart Manufacturing*. London: Academic Press (2019). p. 3–28.
65. Grieves M. (2018). Available at: http://www.apriso.com/library/Whitepaper_Dr_Grieves_DigitalTwin_ManufacturingExcellence.phphttp://www.apriso.com/library/Whitepaper_Dr_Grieves_DigitalTwin_ManufacturingExcellence.php.
66. Rasheed A, San O, Kvamsdal T. Digital Twin: Values, Challenges and Enablers From a Modeling Perspective. *IEEE Access* (2020) 8:21980–2012. doi: 10.1109/ACCESS.2020.2970143
67. Tao F, Qi QL. Make More Digital Twins. *Nature* (2019) 573(7775):490–1. doi: 10.1038/d41586-019-02849-1
68. Chen Y. Integrated and Intelligent Manufacturing: Perspectives and Enablers. *Engineering* (2017) 3(5):588–95. doi: 10.1016/J.ENG.2017.04.009
69. Liu Z, Meyendorf N, Mrad N. The Role of Data Fusion in Predictive Maintenance Using Digital Twin. *AIP Conf Proc* (2018) 1949(1):020023. doi: 10.1063/1.5031520
70. Zheng Y, Yang S, Cheng H. An Application Framework of Digital Twin and its Case Study. *J Ambient Intell Humanized Computing* (2019) 10(3):1141–53. doi: 10.1007/s12652-018-0911-3
71. Vrabič R, Erkoyuncu JA, Butala P, Roy R. Digital Twins: Understanding the Added Value of Integrated Models for Through-Life Engineering Services. *Proc Manufacturing* (2018) 16:139–46. doi: 10.1016/j.promfg.2018.10.167
72. Madni AM, Madni CC, Lucero SD. Leveraging Digital Twin Technology in Model-Based Systems Engineering. *Systems* (2019) 7(1):7. doi: 10.3390/systems7010007
73. Siemens. (2020). Available at: https://new.siemens.com/global/en/company/stories/industry/the-digital-twin.htmlhttps://new.siemens.com/global/en/company/stories/industry/the-digital-twin.html.
74. Cimino C, Negri E, Fumagalli L. Review of Digital Twin Applications in Manufacturing. *Comput Industry* (2019) 113:103130. doi: 10.1016/j.compind.2019.103130
75. Kritzinger W, Karner M, Traar G, Henjes J, Sihn W. Digital Twin in Manufacturing: A Categorical Literature Review and Classification. *IFAC-PapersOnLine* (2018) 51(11):1016–22. doi: 10.1016/j.ifacol.2018.08.474
76. Digital Twin Driven Smart Manufacturing. (2019).
77. Boschert S, Rosen R. Digital Twin—The Simulation Aspect. In: P Hehenberger and D Bradley, editors. *Mechatronic Futures: Challenges and Solutions for Mechatronic Systems and Their Designers*. Cham: Springer International Publishing (2016). p. 59–74.
78. Tao F, Cheng J, Qi Q, Zhang M, Zhang H, Sui F. Digital Twin-Driven Product Design, Manufacturing and Service With Big Data. *Int J Advanced Manufacturing Technol* (2018) 94(9):3563–76. doi: 10.1007/s00170-017-0233-1
79. Urbina Coronado PD, Lynn R, Louhichi W, Parto M, Wescoat E, Kurfess T. Part Data Integration in the Shop Floor Digital Twin: Mobile and Cloud Technologies to Enable a Manufacturing Execution System. *J Manufacturing Syst* (2018) 48:25–33. doi: 10.1016/j.jmsy.2018.02.002
80. Schluse M, Priggemeyer M, Atorf L, Roßmann J. Experimentable Digital Twin" Streamlining Simulation-Based Systems Engineering for Industry 4.0. *IEEE Trans Ind Inf* (2018) 14:1722–31. doi: 10.1109/TII.2018.2804917
81. Laaki H, Miché Y, Tammi K. Prototyping a Digital Twin for Real Time Remote Control Over Mobile Networks: Application of Remote Surgery. *IEEE Access* (2019) 7:20325–36. doi: 10.1109/ACCESS.2019.2897018
82. Jimenez JI, Jahankhani H, Kendzierskyj S. Health Care in the Cyberspace: Medical Cyber-Physical System and Digital Twin Challenges. In: M Farsi, A Daneshkhah, A Hosseinian-Far and H Jahankhani, editors. *Digital Twin Technologies and Smart Cities*. Cham: Springer International Publishing (2020). p. 79–92.
83. Tao F, Zhang H, Liu A, Nee A. Digital Twin in Industry: State-of-the-Art. *IEEE Trans Ind Inf* (2019) 15:2405–15. doi: 10.1109/TII.2018.2873186
84. Kannadasan K, Edla DR, Kuppili V. Type 2 Diabetes Data Classification Using Stacked Autoencoders in Deep Neural Networks. *Clin Epidemiol Global Health* (2019) 7(4):530–5. doi: 10.1016/j.cegh.2018.12.004
85. Schroeder GN, Steinmetz C, Pereira CE, Espindola DB. Digital Twin Data Modeling With AutomationML and a Communication Methodology for Data Exchange. *IFAC-PapersOnLine* (2016) 49(30):12–7. doi: 10.1016/j.ifacol.2016.11.115
86. Haag S, Anderl R. Digital Twin – Proof of Concept. *Manufacturing Lett* (2018) 15:64–6. doi: 10.1016/j.mfglet.2018.02.006
87. Uhlemann THJ, Schock C, Lehmann C, Freiberger S, Steinhilper R. The Digital Twin: Demonstrating the Potential of Real Time Data Acquisition in Production Systems. *Proc Manufacturing* (2017) 9:113–20. doi: 10.1016/j.promfg.2017.04.043
88. Tao F, Zhang M, Liu Y, Nee AYC. Digital Twin Driven Prognostics and Health Management for Complex Equipment. *Cirp Ann-Manuf Techn* (2018) 67:169–72. doi: 10.1016/j.cirp.2018.04.055
89. Hirsch MC. Warum Intelligente Decision-Support-Systeme Das Betriebssystem Eines Smart Hospitals Sein Und Medizin Menschlicher Machen Werden. In: JA Werner, M Forsting, T Kaazte and A Schmidt-Rumposch, editors. *Smart Hospital - Digitale Und Empathische Zukunftsmedizin*. Berlin: MMV Medizinisch Wissenschaftliche Verlagsgesellschaft (2020). p. 93–102.
90. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP. Machine Learning in Cardiovascular Medicine: Are We There Yet? *Heart* (2018) 104 (14):1156–64. doi: 10.1136/heartjnl-2017-311198
91. Rajula HSR, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of Conventional Statistical Methods With Machine Learning in Medicine: Diagnosis, Drug Development, and Treatment. *Med (Kaunas Lithuania)* (2020) 56(9):455. doi: 10.3390/medicina56090455
92. Winter NR, Hahn T. Big Data, AI and Machine Learning for Precision Psychiatry: How are They Changing the Clinical Practice? *Fortschr Der Neurol-Psychiatr* (2020) 88(12):786–93. doi: 10.1055/a-1234-6247
93. Alber M, Buganza Tepole A, Cannon WR, De S, Dura-Bernal S, Garikipati K, et al. Integrating Machine Learning and Multiscale Modeling—Perspectives, Challenges, and Opportunities in the Biological, Biomedical, and Behavioral Sciences. *NPJ Digit Med* (2019) 2(1):115. doi: 10.1038/s41746-019-0193-y
94. Rao D, Mane S. Digital Twin Approach to Clinical DSS With Explainable Ai. (2019) Available at: https://arxiv.org/abs/1910.13520v1.
95. Digital Twins in Healthcare. (2020). Available at: https://www.persistent.com/whitepaper-digital-twins-in-healthcare/.
96. Digitwins. (2018). Available at: https://www.digitwins.orghttps://www.digitwins.org.
97. Filippo MD, Damiani C, Vanoni M, Maspero D, Mauri G, Alberghina L, et al. Single-Cell Digital Twins for Cancer Preclinical Investigation. *Methods Mol Biol (Clifton NJ)* (2020) 2088:331–43. doi: 10.1007/978-1-0716-0159-4_15

98. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-End Lung Cancer Screening With Three-Dimensional Deep Learning on Low-Dose Chest Computed Tomography. *Nat Med* (2019) 25(6):954–61. doi: 10.1038/s41591-019-0447-x
99. Wilhelm D, Berlet M, Feussner H, Ostler D. Digitalisierung in Der Onkologischen Chirurgie. *Forum* (2021) 36:22–8. doi: 10.1007/s12312-020-00879-9
100. Zhang J, Qian H, Zhou H. Application and Research of Digital Twin Technology in Safety and Health Monitoring of the Elderly in Community. *Zhongguo Yi Liao Qi Xie Za Zhi Chin J Med Instrumentation* (2019) 43(6):410–3. doi: 10.3969/j.issn.1671-7104.2019.06.005
101. Calderita LV, Vega A, Barroso-Ramírez S, Bustos P, Núñez P. Designing a Cyber-Physical System for Ambient Assisted Living: A Use-Case Analysis for Social Robot Navigation in Caregiving Centers. *Sensors (Basel Switzerland)* (2020) 20(14):4005. doi: 10.3390/s20144005
102. Hirschvogel M, Jagschies L, Maier A, Wildhirt SM, Gee MW. An in Silico Twin for Epicardial Augmentation of the Failing Heart. *Int J Numerical Methods Biomed Eng* (2019) 35(10):e3233. doi: 10.1002/cnm.3233
103. Hose DR, Lawford PV, Huberts W, Hellevik LR, Omholt SW, van de Vosse FN. Cardiovascular Models for Personalised Medicine: Where Now and Where Next? *Med Eng Phys* (2019) 72:38–48. doi: 10.1016/j.medengphy.2019.08.007
104. Mazumder O, Roy D, Bhattacharya S, Sinha A, Pal A eds. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. (2019) 5024–29. doi: 10.1016/j.medengphy.2019.08.00710.1109/EMBC.2019.8856691
105. Niederer SA, Aboelkassem Y, Cantwell CD, Corrado C, Coveney S, Cherry EM, et al. Creation and Application of Virtual Patient Cohorts of Heart Models. *Philos Trans A Math Phys Eng Sci* (2020) 378:20190558. doi: 10.1098/rsta.2019.0558
106. Sharma P, Suehling M, Flohr T, Comaniciu D. Artificial Intelligence in Diagnostic Imaging: Status Quo, Challenges, and Future Opportunities. *J Thoracic Imaging* (2020) 35 Suppl 1:S11–s6. doi: 10.1097/RTI.0000000000000499
107. Ivanov D. Predicting the Impacts of Epidemic Outbreaks on Global Supply Chains: A Simulation-Based Analysis on the Coronavirus Outbreak (COVID-19/SARS-CoV-2) Case. *Transportation Res Part E Logistics Transport Rev* (2020) 136:101922. doi: 10.1016/j.tre.2020.101922
108. Tellechea-Luzardo J, Winterhalter C, Widera P, Kozyra J, de Lorenzo V, Krasnogor N. Linking Engineered Cells to Their Digital Twins: A Version Control System for Strain Engineering. *ACS Synthetic Biol* (2020) 9(3):536–45. doi: 10.1021/acssynbio.9b00400
109. Lauzeral N, Borzacchiello D, Kugler M, George D, Rémond Y, Hostettler A, et al. A Model Order Reduction Approach to Create Patient-Specific Mechanical Models of Human Liver in Computational Medicine Applications. *Comput Methods Programs Biomed* (2019) 170:95–106. doi: 10.1016/j.cmpb.2019.01.003
110. Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, et al. A Clinically Applicable Approach to Continuous Prediction of Future Acute Kidney Injury. *Nature* (2019) 572(7767):116–9. doi: 10.1038/s41586-019-1390-1
111. Pizzolato C, Saxby DJ, Palipana D, Diamond LE, Barrett RS, Teng YD, et al. Neuromusculoskeletal Modeling-Based Prostheses for Recovery After Spinal Cord Injury. *Front Neurorobot* (2019) 13:97–. doi: 10.3389/fnbot.2019.00097
112. Chakshu NK, Carson J, Sazonov I, Nithiarasu P. A Semi-Active Human Digital Twin Model for Detecting Severity of Carotid Stenoses From Head Vibration-a Coupled Computational Mechanics and Computer Vision Method. *Int J Numerical Methods Biomed Eng* (2019) 35(5):e3180–e. doi: 10.1002/cnm.3180
113. Lareyre F, Adam C, Carrier M, Raffort J. Using Digital Twins for Precision Medicine in Vascular Surgery. *Ann Vasc Surg* (2020) 5096(20):30379-4. doi: 10.1016/j.avsg.2020.04.042
114. Topol EJ. High-Performance Medicine: The Convergence of Human and Artificial Intelligence. *Nat Med* (2019) 25(1):44–56. doi: 10.1038/s41591-018-0300-7
115. The living heart project. (2020). Available at: https://www.3ds.com/products-services/simulia/solutions/life-sciences/the-living-heart-project/https://www.3ds.com/products-services/simulia/solutions/life-sciences/the-living-heart-project/.
116. Grätzel von Grätz P. (2019). Available at: https://www.siemens-healthineers.com/dk/news/mso-solutions-for-individual-patients.htmlhttps://www.siemens-healthineers.com/dk/news/mso-solutions-for-individual-patients.html.
117. Barricelli BR, Casiraghi E, Fogli D. A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications. *IEEE Access* (2019) 7:167653–71. doi: 10.1109/ACCESS.2019.2953499
118. Walsh JR, Smith AM, Pouliot Y, Li-Bland D, Loukianov A, Fisher CK. Generating Digital Twins With Multiple Sclerosis Using Probabilistic Neural Networks. (2020). doi: 10.1101/2020.02.04.934679
119. D Petrova-Antonova, I Spasov, I Krasteva, I Manova and S Ilieva eds. *A Digital Twin Platform for Diagnostics and Rehabilitation of Multiple Sclerosis*. Cham: Springer International Publishing (2020).
120. Fuller A, Fan Z, Day C. Digital Twin: Enabling Technologies, Challenges and Open Research. *IEEE Access* (2020) 8:108952–71. doi: 10.1109/ACCESS.2020.2998358
121. Ziemssen T, Hillert J, Butzkueven H. The Importance of Collecting Structured Clinical Information on Multiple Sclerosis. *BMC Med* (2016) 14:81. doi: 10.1186/s12916-016-0627-1
122. Magyari M, Sorensen PS. Comorbidity in Multiple Sclerosis. *Front Neurol* (2020) 11(851). doi: 10.3389/fneur.2020.00851
123. Toscano S, Patti F. CSF Biomarkers in Multiple Sclerosis: Beyond Neuroinflammation. *Neuroimmunol Neuroinflamm* (2020) 7:14–41. doi: 10.20517/2347-8659.2020.12
124. Ziemssen T, Akgün K, Brück W. Molecular Biomarkers in Multiple Sclerosis. *J Neuroinflamm* (2019) 16(1):272. doi: 10.1186/s12974-019-1674-2
125. Thebault S, Booth RA, Freedman MS. Blood Neurofilament Light Chain: The Neurologist's Troponin? *Biomedicines* (2020) 8(11):523. doi: 10.3390/biomedicines8110523
126. Ziemssen T, Piani-Meier D, Bennett B, Johnson C, Tinsley K, Trigg A, et al. A Physician-Completed Digital Tool for Evaluating Disease Progression (Multiple Sclerosis Progression Discussion Tool): Validation Study. *J Med Internet Res* (2020) 22(2):e16932. doi: 10.2196/16932
127. D'Souza M, Yaldizli Ö, John R, Vogt DR, Papadopoulou A, Lucassen E, et al. Neurostatus e-Scoring Improves Consistency of Expanded Disability Status Scale Assessments: A Proof of Concept Study. *Mult Scler* (2017) 23(4):597–603. doi: 10.1177/1352458516657439
128. Kosa P, Barbour C, Wichman A, Sandford M, Greenwood M, Bielekova B. NeurEx: Digitalized Neurological Examination Offers a Novel High-Resolution Disability Scale. *Ann Clin Trans Neurol* (2018) 5(10):1241–9. doi: 10.1002/acn3.640
129. Kurtzke JFM. Rating Neurologic Impairment in Multiple Sclerosis: An Expanded Disability Status Scale (EDSS). *Neurology* (1983) 33(11):1444–52. doi: 10.1212/WNL.33.11.1444
130. Beste C, Mückschel M, Paucke M, Ziemssen T. Dual-Tasking in Multiple Sclerosis - Implications for a Cognitive Screening Instrument. *Front Hum Neurosci* (2018) 12:24. doi: 10.3389/fnhum.2018.00024
131. Trentzsch K, Weidemann ML, Torp C, Inojosa H, Scholz M, Haase R, et al. The Dresden Protocol for Multidimensional Walking Assessment (DMWA) in Clinical Practice. *Front Neurosci* (2020) 14:582046. doi: 10.3389/fnins.2020.582046
132. Lublin FD, Reingold SC, Cohen JA, Cutter GR, Sørensen PS, Thompson AJ, et al. Defining the Clinical Course of Multiple Sclerosis. *2013 Revisions* (2014) 83(3):278–86. doi: 10.1212/wnl.46.4.907
133. De Meo E, Portaccio E, Giorgio A, Ruano L, Goretti B, Niccolai C, et al. Identifying the Distinct Cognitive Phenotypes in Multiple Sclerosis. *JAMA Neurol* (2021)78(4):414–425. doi: 10.1001/jamaneurol.2020.4920
134. Inojosa H, Schriefer D, Ziemssen T. Clinical Outcome Measures in Multiple Sclerosis: A Review. *Autoimmun Rev* (2020) 19(5):102512. doi: 10.1016/j.autrev.2020.102512
135. D'Souza M, Papadopoulou A, Girardey C, Kappos L. Standardization and Digitization of Clinical Data in Multiple Sclerosis. *Nat Rev Neurol* (2021):119–125. doi: 10.1038/s41582-020-00448-7
136. Meyer zu Hörste G, Gross CC, Klotz L, Schwab N, Wiendl H. Next-Generation Neuroimmunology: New Technologies to Understand Central Nervous System Autoimmunity. *Trends Immunol* (2020)41(4):341–354. doi: 10.1016/j.it.2020.02.005

137. Leocani L, Rocca MA, Comi G. MRI and Neurophysiological Measures to Predict Course, Disability and Treatment Response in Multiple Sclerosis. *Curr Opin Neurol* (2016) 29(3):243–53. doi: 10.1097/WCO.0000000000000333

138. Marciniewicz E, Podgórski P, Sąsiadek M, Bladowska J. The Role of MR Volumetry in Brain Atrophy Assessment in Multiple Sclerosis: A Review of the Literature. *Adv Clin Exp Med Off Organ Wroclaw Med Univ* (2019) 28 (7):989–99. doi: 10.17219/acem/94137

139. Louapre C. Conventional and Advanced MRI in Multiple Sclerosis. *Rev Neurol* (2018) 174(6):391–7. doi: 10.1016/j.neurol.2018.03.009

140. Kaufmann M, Haase R, Proschmann U, Ziemssen T, Akgün K. Real-World Lab Data in Natalizumab Treated Multiple Sclerosis Patients Up to 6 Years Long-Term Follow Up. *Front Neurol* (2018) 9:1071. doi: 10.3389/fneur.2018.01071

141. Kaufmann M, Haase R, Proschmann U, Ziemssen T, Akgün K. Real World Lab Data: Patterns of Lymphocyte Counts in Fingolimod Treated Patients. *Front Immunol* (2018) 9:2669. doi: 10.3389/fimmu.2018.02669

142. Barro C, Chitnis T, Weiner HL. Blood Neurofilament Light: A Critical Review of its Application to Neurologic Disease. *Ann Clin Trans Neurol* (2020) 7(12):2508–23. doi: 10.1002/acn3.51234

143. Akgün K, Kretschmann N, Haase R, Proschmann U, Kitzler HH, Reichmann H, et al. Profiling Individual Clinical Responses by High-Frequency Serum Neurofilament Assessment in MS. *Neurol(R) Neuroimmunol Neuroinflamm* (2019) 6(3):e555. doi: 10.1212/NXI.0000000000000555

144. Cortese R, Collorone S, Ciccarelli O, Toosy AT. Advances in Brain Imaging in Multiple Sclerosis. *Ther Adv Neurol Disord* (2019) 121–15. doi: 10.1177/1756286419859722

145. Oreja-Guevara C. Overview of Magnetic Resonance Imaging for Management of Relapsing-Remitting Multiple Sclerosis in Everyday Practice. *Eur J Neurol* (2015) 22 Suppl 2:22–7. doi: 10.1111/ene.12800

146. Tomassini V, Sinclair A, Sawlani V, Overell J, Pearson OR, Hall J, et al. Diagnosis and Management of Multiple Sclerosis: MRI in Clinical Practice. *J Neurol* (2020) 267(10):2917–25. doi: 10.1007/s00415-020-09930-0

147. Oh J, Sicotte NL. New Imaging Approaches for Precision Diagnosis and Disease Staging of MS? *Mult Scler* (2020) 26(5):568–75. doi: 10.1177/1352458519871817

148. Arevalo O, Riascos R, Rabiei P, Kamali A, Nelson F. Standardizing Magnetic Resonance Imaging Protocols, Requisitions, and Reports in Multiple Sclerosis: An Update for Radiologist Based on 2017 Magnetic Resonance Imaging in Multiple Sclerosis and 2018 Consortium of Multiple Sclerosis Centers Consensus Guidelines. *J Comput Assisted Tomography* (2019) 43 (1):1–12. doi: 10.1097/rct.0000000000000767

149. Saslow L, Li DKB, Halper J, Banwell B, Barkhof F, Barlow L, et al. An International Standardized Magnetic Resonance Imaging Protocol for Diagnosis and Follow-up of Patients With Multiple Sclerosis: Advocacy, Dissemination, and Implementation Strategies. *Int J MS Care* (2020) 22 (5):226–32. doi: 10.7224/1537-2073.2020-094

150. Pessini RA, ACd S, Salmon CEG. Quantitative MRI Data in Multiple Sclerosis Patients: A Pattern Recognition Study. *Res Biomed Eng* (2018) 34:138–46. doi: 10.1590/2446-4740.07117

151. Afzal HMR, Luo S, Ramadan S, Lechner-Scott J. The Emerging Role of Artificial Intelligence in Multiple Sclerosis Imaging. *Mult Scler* (2020) 1–10. doi: 10.1177/1352458520966298

152. Tauhid S, Neema M, Healy BC, Weiner HL, Bakshi R. MRI Phenotypes Based on Cerebral Lesions and Atrophy in Patients With Multiple Sclerosis. *J Neurol Sci* (2014) 346(1-2):250–4. doi: 10.1016/j.jns.2014.08.047

153. Hanson JVM, Wicki CA, Manogaran P, Petzold A, Schippling S. OCT and Imaging in Central Nervous System Diseases: The Eye as a Window to the Brain. In: *OCT and Multiple Sclerosis*, *2nd ed.* (2020). p. 195–233. doi: 10.1007/978-3-030-26269-3_11

154. Bauckneht M, Capitanio S, Raffa S, Roccatagliata L, Pardini M, Lapucci C, et al. Molecular Imaging of Multiple Sclerosis: From the Clinical Demand to Novel Radiotracers. *EJNMMI Radiopharmacy Chem* (2019) 4(1):6. doi: 10.1186/s41181-019-0058-3

155. Dorsey ER, Papapetropoulos S, Xiong M, Kieburtz K. The First Frontier: Digital Biomarkers for Neurodegenerative Disorders. *Digital Biomarkers* (2017) 1(1):6–13. doi: 10.1159/000477383

156. Perry B, Herrington W, Goldsack JC, Grandinetti CA, Vasisht KP, Landray MJ, et al. Use of Mobile Devices to Measure Outcomes in Clinical Research, 2010-2016: A Systematic Literature Review. *Digital Biomarkers* (2018) 2 (1):11–30. doi: 10.1159/000486347

157. Kang C, Janes H, Tajik P, Groen H, Mol B, Koopmans C, et al. Evaluation of Biomarkers for Treatment Selection Using Individual Participant Data From Multiple Clinical Trials. *Stat Med* (2018) 37(9):1439–53. doi: 10.1002/sim.7608

158. Dagum P. Digital Biomarkers of Cognitive Function. *NPJ Digit Med* (2018) 1 (1):10. doi: 10.1038/s41746-018-0018-4

159. Barrios L, Oldrati P, Santini S, Lutterotti A. (2018). Recognizing Digital Biomarkers for Fatigue Assessment in Patients with Multiple Sclerosis.

160. Rudick RA, Miller D, Bethoux F, Rao SM, Lee JC, Stough D, et al. The Multiple Sclerosis Performance Test (MSPT): An iPad-based Disability Assessment Tool. *J Vis Exp* (2014) 88):e51318. doi: 10.3791/51318

161. Rao SM, Galioto R, Sokolowski M, McGinley M, Freiburger J, Weber M, et al. Multiple Sclerosis Performance Test: Validation of Self-Administered Neuroperformance Modules. *Eur J Neurol* (2020) 27(5):878–86. doi: 10.1111/ene.14162

162. Rao SM, Losinski G, Mourany L, Schindler D, Mamone B, Reece C, et al. Processing Speed Test: Validation of a Self-Administered, iPad(®)-based Tool for Screening Cognitive Dysfunction in a Clinic Setting. *Mult Scler* (2017) 23(14):1929–37. doi: 10.1177/1352458516688955

163. Gijbels D, Eijnde BO, Feys P. Comparison of the 2- and 6-Minute Walk Test in Multiple Sclerosis. *Mult Scler* (2011) 17(10):1269–72. doi: 10.1177/1352458511408475

164. Rossier P, Wade DT. Validity and Reliability Comparison of 4 Mobility Measures in Patients Presenting With Neurologic Impairment. *Arch Phys Med Rehabil* (2001) 82(1):9–13. doi: 10.1053/apmr.2001.9396

165. Marziniak M, Brichetto G, Feys P, Meyding-Lamade U, Vernon K, Meuth SG. The Use of Digital and Remote Communication Technologies as a Tool for Multiple Sclerosis Management: Narrative Review. *JMIR Rehabil Assist Technol* (2018) 5(1):e5. doi: 10.2196/rehab.7805

166. Scholz M, Haase R, Schriefer D, Voigt I, Ziemssen T. Electronic Health Interventions in the Case of Multiple Sclerosis: From Theory to Practice. *Brain Sci* (2021) 11(2):180. doi: 10.3390/brainsci11020180

167. Babre D. Clinical Data Interchange Standards Consortium: A Bridge to Overcome Data Standardisation. *Perspect Clin Res* (2013) 4(2):115–6. doi: 10.4103/2229-3485.111779

168. Peeters LM, Parciak T, Kalra D, Moreau Y, Kasilingam E, van Galen P, et al. Multiple Sclerosis Data Alliance – A Global Multi-Stakeholder Collaboration to Scale-Up Real World Data Research. *Multiple Sclerosis Related Disord* (2021) 47:102634. doi: 10.1016/j.msard.2020.102634

169. Tilocca B, Pieroni L, Soggiu A, Britti D, Bonizzi L, Roncada P, et al. Gut-Brain Axis and Neurodegeneration: State-of-the-Art of Meta-Omics Sciences for Microbiota Characterization. *Int J Mol Sci* (2020) 21(11):4045. doi: 10.3390/ijms21114045

170. Martin NA, Nawrocki A, Molnar V, Elkjaer ML, Thygesen EK, Palkovits M, et al. Orthologous Proteins of Experimental De- and Remyelination are Differentially Regulated in the CSF Proteome of Multiple Sclerosis Subtypes. *PloS One* (2018) 13(8):e0202530. doi: 10.1371/journal.pone.0202530

171. Malekzadeh A, Teunissen C. Recent Progress in Omics-Driven Analysis of MS to Unravel Pathological Mechanisms. *Expert Rev Neurother* (2013) 13 (9):1001–16. doi: 10.1586/14737175.2013.835602

172. Sun YV, Hu Y-J. Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases. *Adv Genet* (2016) 93:147–90. doi: 10.1016/bs.adgen.2015.11.004

173. Chase Huizar C, Raphael I, Forsthuber TG. Genomic, Proteomic, and Systems Biology Approaches in Biomarker Discovery for Multiple Sclerosis. *Cell Immunol* (2020) 358:104219. doi: 10.1016/j.cellimm.2020.104219

174. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-Omics Data Integration, Interpretation, and Its Application. *Bioinf Biol Insights* (2020) 14:1177932219899051. doi: 10.1177/1177932219899051

175. Klose K, Kreimeier S, Tangermann U, Aumann I, Damm Kon behalf of the RHOG. Patient- and Person-Reports on Healthcare: Preferences, Outcomes, Experiences, and Satisfaction – an Essay. *Health Econom Rev* (2016) 6(1):18. doi: 10.1186/s13561-016-0094-6

176. D'Amico E, Haase R, Ziemssen T. Review: Patient-reported Outcomes in Multiple Sclerosis Care. *Mult Scler Relat Disord* (2019) 33:61–6. doi: 10.1016/j.msard.2019.05.019

177. Medina LD, Torres S, Alvarez E, Valdez B, Nair KV. Patient-Reported Outcomes in Multiple Sclerosis: Validation of the Quality of Life in

Neurological Disorders (Neuro-QoL^{TM}) Short Forms. *Mult Scler J Exp Transl Clin* (2019) 5(4):1–11. doi: 10.1177/2055217319885986

178. Cella D, Lai JS, Nowinski CJ, Victorson D, Peterman A, Miller D, et al. Neuro-QOL: Brief Measures of Health-Related Quality of Life for Clinical Research in Neurology. *Neurology* (2012) 78(23):1860–7. doi: 10.1212/WNL.0b013e318258f744
179. Hobart JC, Riazi A, Lamping DL, Fitzpatrick R, Thompson AJ. Measuring the Impact of MS on Walking Ability: The 12-Item MS Walking Scale (Msws-12). *Neurology* (2003) 60(1):31–6. doi: 10.1212/wnl.60.1.31
180. Ziemssen T, Phillips G, Shah R, Mathias A, Foley C, Coon C, et al. Development of the Multiple Sclerosis (MS) Early Mobility Impairment Questionnaire (EMIQ). *J Neurol* (2016) 263(10):1969–83. doi: 10.1007/s00415-016-8210-4
181. Hodson M, Andrew S, Michael Roberts C. Towards an Understanding of PREMS and PROMS in COPD. *Breathe* (2013) 9(5):358–64. doi: 10.1183/20734735.006813
182. Male L, Noble A, Atkinson J, Marson T. Measuring Patient Experience: A Systematic Review to Evaluate Psychometric Properties of Patient Reported Experience Measures (Prems) for Emergency Care Service Provision. *Int J Qual Health Care* (2017) 29(3):314–26. doi: 10.1093/intqhc/mzx027
183. The Lancet N. Patient-Reported Outcomes in the Spotlight. *Lancet Neurol* (2019) 18(11):981. doi: 10.1016/S1474-4422(19)30357-6
184. Wiendl H, Korsukewitz C, Kieseier BC. Klinik, Diagnostik Und Therapie. Klinische Neurologie. In: *Multiple Sklerose*. Stuttgart: Kohlhammer (2021).
185. Giovannoni G, Bermel R, Phillips T, Rudick R. A Brief History of NEDA. *Mult Scler Relat Disord* (2018) 20:228–30. doi: 10.1016/j.msard.2017.07.011
186. Giovannoni G, Butzkueven HD-JS, Hobart J, Kobelt G, Pepper G, Sormani MP, et al. (2018). Brain Health: Keine Zeit verlieren bei Multipler Sklerose.
187. Linker R, Kallmann BA, Kleinschnitz C, Rieckmann P, Maurer M, Schwab S. "Time is Brain" in Relapsing Remitting Multiple Sclerosis. *Curr Treat Concepts Immunother Nervenarzt* (2015) 86(12):1528–37. doi: 10.1007/s00115-015-4439-x
188. Soelberg Sorensen P, Giovannoni G, Montalban X, Thalheim C, Zaratin P, Comi G. The Multiple Sclerosis Care Unit. *Mult Scler* (2018) 418:627–36. doi: 10.1177/1352458518807082
189. Stühler E, Braune S, Lionetto F, Heer Y, Jules E, Westermann C, et al. Framework for Personalized Prediction of Treatment Response in Relapsing Remitting Multiple Sclerosis. *BMC Med Res Methodol* (2020) 20(1):24. doi: 10.1186/s12874-020-0906-6
190. Kalincik T, Manouchehrinia A, Sobisek L, Jokubaitis V, Spelman T, Horakova D, et al. Towards Personalized Therapy for Multiple Sclerosis: Prediction of Individual Treatment Response. *Brain* (2017) 140(9):2426–43. doi: 10.1093/brain/awx185
191. Kalincik T. Reply: Towards Personalized Therapy for Multiple Sclerosis: Limitations of Observational Data. *Brain* (2018)141(5):e39. doi: 10.1093/brain/awy056
192. Pinto MF, Oliveira H, Batista S, Cruz L, Pinto M, Correia I, et al. Prediction of Disease Progression and Outcomes in Multiple Sclerosis With Machine Learning. *Sci Rep* (2020) 10(1):21038. doi: 10.1038/s41598-020-78212-6
193. HL7 FHIR. (2017). Available at: http://hl7.org/fhir/http://hl7.org/fhir/.
194. XDS Value Sets für Deutschland. (2018). Available at: http://www.ihe-d.de/projekte/xds-value-sets-fuer-deutschland/http://www.ihe-d.de/projekte/xds-value-sets-fuer-deutschland/.
195. Voigt I, Benedict M, Susky M, Scheplitz T, Frankowitz S, Kern R, et al. A Digital Patient Portal for Patients With Multiple Sclerosis. *Front Neurol* (2020) 11(400). doi: 10.3389/fneur.2020.00400
196. Benedict M, Schlieter H, Burwitz M, Scheplitz T, Susky M, Richter P, et al. Patientenintegration Durch Pfadsysteme. *Wirtschaftsinformatik* (2019). Siegen. Available at: https://www.researchgate.net/publication/330203014_Patientenintegration_durch_Pfadsysteme.
197. Benedict P, Schlieter H. Understanding Patient Pathways in the Context of Integrated Health Care Services - Implications From a Scoping Review. *14 Internationalen Tagung Wirtschaftsinformatik* (2019). Siegen. Available at: https://www.semanticscholar.org/paper/Understanding-Patient-Pathways-in-the-Context-of-a-Richter-Schlieter/dca3766d0c79a152da2f62421caef03ce455e96c.
198. Minkman M, Ahaus K, Fabbricotti I, Nabitz U, Huijsman R. A Quality Management Model for Integrated Care: Results of a Delphi and Concept Mapping Study. *Int J Qual Health Care* (2009) 21(1):66–75. doi: 10.1093/intqhc/mzn048
199. Multiple sclerosis. (2016). Available at: https://www.nice.org.uk/guidance/qs108https://www.nice.org.uk/guidance/qs108.
200. Patient experience in adult NHS services. (2012). Available at: https://www.nice.org.uk/guidance/qs15https://www.nice.org.uk/guidance/qs15.
201. Haase R, Wunderlich M, Dillenseger A, Kern R, Akgun K, Ziemssen T. Improving Multiple Sclerosis Management and Collecting Safety Information in the Real World: The MSDS3D Software Approach. *Expert Opin Drug Saf* (2018) 17(4):369–78. doi: 10.1080/14740338.2018.1437144
202. Ziemssen T, Kern R, Voigt I, Haase R. Data Collection in Multiple Sclerosis: The Msds Approach. *Front Neurol* (2020) 11(445). doi: 10.3389/fneur.2020.00445
203. Schleimer E, Pearce J, Barnecut A, Rowles W, Lizee A, Klein A, et al. A Precision Medicine Tool for Patients With Multiple Sclerosis (the Open Ms BioScreen): Human-Centered Design and Development. *J Med Internet Res* (2020) 22(7):e15605. doi: 10.2196/15605
204. Shortliffe EH, Sepúlveda MJ. Clinical Decision Support in the Era of Artificial Intelligence. *Jama* (2018) 320(21):2199–200. doi: 10.1001/jama.2018.17163
205. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An Overview of Clinical Decision Support Systems: Benefits, Risks, and Strategies for Success. *NPJ Digit Med* (2020) 3:17. doi: 10.1038/s41746-020-0221-y
206. Alshamrani R, Althbiti A, Alshamrani Y, Alkomah F, Ma X. Model-Driven Decision Making in Multiple Sclerosis Research: Existing Works and Latest Trends. *Patterns (N Y NY)* (2020) 1(8):100121. doi: 10.1016/j.patter.2020.100121
207. Zhao Y, Healy BC, Rotstein D, Guttmann CR, Bakshi R, Weiner HL, et al. Exploration of Machine Learning Techniques in Predicting Multiple Sclerosis Disease Course. *PloS One* (2017) 12(4):e0174866. doi: 10.1371/journal.pone.0174866
208. Zhao Y, Wang T, Bove R, Cree B, Henry R, Lokhande H, et al. Ensemble Learning Predicts Multiple Sclerosis Disease Course in the SUMMIT Study. *NPJ Digit Med* (2020) 3:135. doi: 10.1038/s41746-020-00361-9
209. Goyal M, Khanna D, Rana PS, Khaibullin T, Martynova E, Rizvanov AA, et al. Computational Intelligence Technique for Prediction of Multiple Sclerosis Based on Serum Cytokines. *Front Neurol* (2019) 10:781. doi: 10.3389/fneur.2019.00781
210. Deans KJ, Sabihi S, Forrest CB. Learning Health Systems. *Semin Pediatr Surg* (2018) 27(6):375–8. doi: 10.1053/j.sempedsurg.2018.10.005
211. Warraich HJ, Califf RM, Krumholz HM. The Digital Transformation of Medicine can Revitalize the Patient-Clinician Relationship. *NPJ Digit Med* (2018) 1:49. doi: 10.1038/s41746-018-0060-2
212. Verghese A, Shah NH, Harrington RA. What This Computer Needs is a Physician: Humanism and Artificial Intelligence. *JAMA* (2018) 319(1):19–20. doi: 10.1001/jama.2017.19198
213. Nordling L. A Fairer Way Forward for AI in Health Care. *Nature* (2019) 573 (7775):S103–s5. doi: 10.1038/d41586-019-02872-2
214. Asan O, Bayrak AE, Choudhury A. Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. *J Med Internet Res* (2020) 22(6):e15154. doi: 10.2196/15154
215. Bhattad PB, Jain V. Artificial Intelligence in Modern Medicine - The Evolving Necessity of the Present and Role in Transforming the Future of Medical Care. *Cureus* (2020) 12(5):e8041. doi: 10.2196/preprints.18829

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

ORIGINAL RESEARCH
published: 31 May 2021
doi: 10.3389/fimmu.2021.683330

Single-Cell Profiling Reveals Transcriptional Signatures and Cell-Cell Crosstalk in Anti-PLA2R Positive Idiopathic Membranous Nephropathy Patients

Jie Xu[1], Chanjuan Shen[2], Wei Lin[3], Ting Meng[1], Joshua D. Ooi[1,4], Peter J. Eggenhuizen[4], Rong Tang[1], Gong Xiao[1], Peng Jin[5], Xiang Ding[5], Yangshuo Tang[6], Weisheng Peng[1], Wannian Nie[1], Xiang Ao[1], Xiangcheng Xiao[1], Yong Zhong[1] and Qiaoling Zhou[1]**

[1] Department of Nephrology, Xiangya Hospital, Central South University, Changsha, China, [2] Department of Hematology, The Affiliated Zhuzhou Hospital Xiangya Medical College, Central South University, Zhuzhou, China, [3] Department of Pathology, Xiangya Hospital, Central South University, Changsha, China, [4] Centre for Inflammatory Diseases, Monash University Department of Medicine, Monash Medical Centre, Clayton, VIC, Australia, [5] Department of Organ Transplantation, Xiangya Hospital, Central South University, Changsha, China, [6] Department of Ultrasonography, Xiangya Hospital, Central South University, Changsha, China

OPEN ACCESS

Edited by:
Xu-jie Zhou,
Peking University First Hospital, China

Reviewed by:
Chia-Chao Wu,
Tri-Service General Hospital, Taiwan
Jianfang Cai,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

***Correspondence:**
Yong Zhong
zhongyong121@163.com
Qiaoling Zhou
zqling8315@163.com

Specialty section:
This article was submitted to Autoimmune and Autoinflammatory Disorders, a section of the journal Frontiers in Immunology

Received: 20 March 2021
Accepted: 11 May 2021
Published: 31 May 2021

Citation:
Xu J, Shen C, Lin W, Meng T, Ooi JD, Eggenhuizen PJ, Tang R, Xiao G, Jin P, Ding X, Tang Y, Peng W, Nie W, Ao X, Xiao X, Zhong Y and Zhou Q (2021) Single-Cell Profiling Reveals Transcriptional Signatures and Cell-Cell Crosstalk in Anti-PLA2R Positive Idiopathic Membranous Nephropathy Patients. Front. Immunol. 12:683330. doi: 10.3389/fimmu.2021.683330

Idiopathic membranous nephropathy (IMN) is an organ-specific autoimmune disease of the kidney glomerulus. It may gradually progress to end-stage renal disease (ESRD) characterized by increased proteinuria, which leads to serious consequences. Although substantial advances have been made in the understanding of the molecular bases of IMN in the last 10 years, certain questions remain largely unanswered. To define the transcriptomic landscape at single-cell resolution, we analyzed kidney samples from 6 patients with anti-PLA2R positive IMN and 2 healthy control subjects using single-cell RNA sequencing. We then identified distinct cell clusters through unsupervised clustering analysis of kidney specimens. Identification of the differentially expressed genes (DEGs) and enrichment analysis as well as the interaction between cells were also performed. Based on transcriptional expression patterns, we identified all previously described cell types in the kidney. The DEGs in most kidney parenchymal cells were primarily enriched in genes involved in the regulation of inflammation and immune response including IL-17 signaling, TNF signaling, NOD-like receptor signaling, and MAPK signaling. Moreover, cell-cell crosstalk highlighted the extensive communication of mesangial cells, which infers great importance in IMN. IMN with massive proteinuria displayed elevated expression of genes participating in inflammatory signaling pathways that may be involved in the pathogenesis of the progression of IMN. Overall, we applied single-cell RNA sequencing to IMN to uncover intercellular interactions, elucidate key pathways underlying the pathogenesis, and identify novel therapeutic targets of anti-PLA2R positive IMN.

Keywords: single-cell RNA sequence, idiopathic membranous nephropathy, kidney, immune response, inflammation

INTRODUCTION

Idiopathic membranous nephropathy (IMN) is a common cause of nephrotic syndrome (NS) in adults with a peak occurrence of 50-60 years old (1). It is characterized by subepithelial immune deposits, complement-mediated proteinuria, and risk of kidney failure. The prevalence of IMN is increasing worldwide, particularly in elderly patients, and has been reported in 20.0–36.8% of adult-onset NS cases (2–5). The clinical outcome of patients is quite variable, with spontaneous remission reported in up to one-third of cases and progression to end-stage renal disease (ESRD) in a similar number (6–8).

IMN is a noninflammatory autoimmune disease of the kidney glomerulus (9, 10). In the last 10 years, substantial advances have been made in the understanding of the molecular bases of IMN, with the identification of several antigens [neutral endopeptidase, phospholipase A2 receptor (PLA2R), thrombospondin domain-containing 7A (THSD7A)] and the characterization of antibody-binding domains of these auto-antigens. 50% to 80% of the patients will test positive for an anti-PLA2R antibody with any of the available tests depending on the state of disease activity (11, 12). These ground-breaking findings already have a major impact on diagnosis and therapy monitoring. Besides, several risk alleles, such as HLA-DQ, HLA-DR, and PLA2R1 have been identified as risk factors of IMN (13, 14). The pathogenesis of IMN induced by podocyte *in situ* antibody and the following complement activation pathways have been revealed to some extent (9, 15, 16). However, the reason for the heterogeneity of patients as well as the variety of clinical outcomes remains elusive. Furthermore, a comprehensive analysis of the cell types and molecular pathways involved in IMN is lacking.

Single-cell RNA-sequencing (scRNA-seq) is a transcriptomic technology that measures the expression of up to thousands of genes in thousands of single cells simultaneously. It offers an opportunity to comprehensively describe human kidney disease at a cellular level and plays a crucial role in identifying cell subtypes and illustrating molecular differences (17). This technique has been applied to several complex kidney diseases including kidney cell carcinoma, diabetic nephropathy, lupus nephritis, and acute kidney injury (18–21). Here we applied scRNA-seq to kidney biopsies of patients with IMN to identify gene expression at the single-cell level, elucidate cells involved in the progression of IMN, and uncover intercellular interactions.

MATERIALS AND METHODS

Ethical Approval and Consent

The Medical Ethics Committee of the Xiangya Hospital of Central South University for Human Studies approved the study (ID: 201711836). The implementations were in concordance with the International Ethical Guidelines for Research Involving Human Subjects as stated in the Declaration of Helsinki. Informed written consent was obtained from participants or their legal guardians.

Tissue Procurement

Kidney specimens were obtained from the department of nephrology in Xiangya Hospital, Central South University. We conducted a kidney biopsy with 18-gauge core needles in the nephrotic syndrome subjects paralleling with positive serum anti-PLA2R antibody. Healthy adult kidney tissues were collected by biopsy of living donor kidneys from two transplant donors. Healthy kidney tissue was collected after removal from the donor and before implantation into the recipient. Kidney tissues were cleaned with sterile phosphate buffered saline (PBS) after collection.

Kidney Sample Processing and Single-Cell Dissociation

Fresh kidney tissue specimens were stored in GEXSCOPE Tissue Preservation Solution (Singleron Biotechnologies) at 2-8°C immediately. The specimens were washed with Hanks' Balanced Salt Solution (HBSS) three times and then minced into 1-2 mm pieces before dissociation. Single-cell suspensions were obtained by digestion with 2ml GEXSCOPE Tissue Dissociation Solution (Singleron Biotechnologies) with continuous agitation at 37°C for 15min. The samples were subsequently filtered through 40-µm sterile cell strainers (Corning) to separate cells from cell debris and other impurities, after which they were centrifuged at 300 x g for 5 minutes at 4°C and cell pellets were resuspended into 1ml PBS (HyClone). Next, 2ml GEXSCOPE Red Blood Cell Lysis Buffer (Singleron Biotechnologies) was added into the cell suspension and incubated at 25°C for 10 minutes to remove red blood cells. The cells were then centrifuged at 300 x g for 5 min and resuspended in cold PBS for downstream analyses. Quantification of cell yields was performed by TC20 automated cell counter (Bio-Rad) with trypan blue exclusion, once the cell viability exceeded 70%, subsequent sample processing could be performed.

Library Preparation and Preprocessing of scRNA-Seq Data

PBS was added to the single-cell suspension to adjust the concentration to 1×10^5 cells/mL. A single-cell suspension was then loaded onto the microfluidic chip. The single-cell RNA-seq libraries were prepared according to the manufacturer's protocol using the Singleron GEXSCOPE Single Cell RNA-seq Library Kit (Singleron Biotechnologies), which included cell lysis, mRNA trapping, labeling cells (barcode) and mRNA (UMI), reverse transcription mRNA into cDNA and amplification, and finally fragment cDNA. Samples were then sequenced by Hiseq X10 (Illumina, San Diego, CA, USA) with 150bp paired-end reads. Raw reads were processed to generate gene expression profiles using an internal pipeline. Adapters and poly-A tails were trimmed (fastp V1) before aligning read two to GRCh38 with ensemble version 92 gene annotation (fastp 2.5.3a and featureCounts 1.6.2). Reads with the same cell barcode, UMI, and gene were grouped to calculate the number of UMIs per gene per cell.

Cell Type Classification and Marker Genes Analysis

The Seurat program (http://satijalab.org/seurat/, R package, v.3.0.1) was applied for the analysis of RNA-Sequencing data including cell type identification and clustering analysis. By default, we used the SNN (shared nearest neighbor) model of the Seurat program package for clustering analyses and displayed the distribution status of cells by dimension reduction operation (PCA, tSNE, UMAP). Next, Wilcox (Wilcoxon rank-sum test) was used to analyze the difference of each cluster and the result of the marker gene obtained by "Wilcox" (Likelihood-ratio test) using the FindAllMarkers function in Seurat combined the differential gene list above identified the marker gene of each cluster. The selected marker genes were expressed in over 10% of the cells per cluster and the average log (Fold Change) was more than 0.25. The heatmap was completed by the top20 marker genes of each cell cluster. Sub-clustering analysis of endothelial cells and myeloid immune cells was performed by the SubsetData function of the Seurat.

Differentially Expressed Genes Identified Between Groups

Differentially expressed genes (DEGs) of each kidney cell cluster were identified by comparing the transcriptional profile of IMN and healthy donors. We performed "Wilcox"(Likelihood-ratio test)by the FindAllMarkers function in Seurat to determine the DEGs of each cluster between the two groups. DEGs were defined by a gene with an average log (Fold Change) exceeded 0.25 and P-value smaller than 0.05.

Enrichment and Cell Interaction Analysis

Gene Ontology (GO) function enrichment analysis was performed on the gene set using the clusterProfiler software to find biological processes or molecular functions that are significantly associated with the genes specifically expressed. Similarly, Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was carried out to get significantly related pathways by the clusterProfiler software. We also conducted a ligand-receptor interaction analysis of cell-cell cross-talk by CellphoneDB.

RESULTS

Extensive clinical, laboratory, and pathologic evaluation were required to separate primary from secondary MN and help determine the underlying etiology (22). All patients were reviewed for potential secondary causes of MN such as hepatitis serology, antinuclear antibodies, anti–double-stranded DNA antibodies, anti–Smith antibodies, complements, chest radiographs, age-appropriate cancer screening, medication history, and monoclonal gammopathy evaluation. All biopsies were reviewed for histologic features suggestive of secondary MN, such as full house immunofluorescence, vascular or tubular basement membrane deposits on immunofluorescence, and mesangial and endothelial proliferation, as well as for the presence of endothelial tubuloreticular inclusions or mesangial deposits on electron microscopy (23). Patients with potential secondary causes were excluded from further analysis. Kidney biopsy samples were obtained from 6 patients with IMN and 2 healthy controls. Patients with IMN were positive for serum anti-PLA2R antibody and evidenced by the diffuse formation of subepithelial "spikes", or heterogeneous thickening of the glomerular basement membrane by light microscopy, as well as subepithelial electron-dense deposits and diffuse fusion of podocyte foot processes by electron microscopy (**Supplemental Figure 1**). Serum anti-PLA2R antibody levels ranged from 26.28 to 816.47 RU/ml and the pathologic stage varied from II-IV. Also, patient ages ranged from 34 to 65 years and one of them was female. The proteinuria of these six IMN patients varied from 1.18 to 11.35 g per 24h. IMN patients were divided into massive proteinuria group and non-massive proteinuria group with the critical value of 3.5g urinary protein excretion per day. Four of the six IMN patients had more than 3.5g proteinuria per 24h in our study. Albumin (ALB) and triglyceride (TG) were 22.22 ± 1.00 g/L and 1.96 ± 0.42 mmol/l respectively. The estimated glomerular filtration rate (eGFR) of IMN patients ranged from 61.91 to 107 mL/min/1.73m^2, and the mean serum creatinine (Scr) was 1.08 ± 0.11 mg/dL. Serum C3 concentration of all subjects was normal whereas serum IgG levels in four IMN patients were below the lower limit of normal (**Supplemental Table 1**). At the time of biopsy, all IMN patients have not received medications other than RAAS inhibitors. As for healthy control, one kidney donor was a 47-year-old male, and his creatinine was 0.68 mg/dl, eGFR was 113.81 mL/min/1.73m^2. Another kidney donor was a 50-year-old male, and his creatinine was 0.94 mg/dl, eGFR was 94.16 mL/min/1.73m^2. The two donors neither have diseases history including hypertension and diabetes nor any medications prescribed before.

Cell Lineage in the Kidney Identified by scRNA-Seq

We first catalogued kidney cell types of all eight subjects in an unbiased manner using droplet-based single-cell RNA sequencing. After data pre-processing and stringent quality control, transcriptomic data were obtained from 30313 cells (**Figure 1A**). The number of cells for each sample varied from 2047 to 8879 and cell viability ranged from 71% to 98% (**Supplemental Table 2**). Eleven kidney subsets and six immune subsets consisting of as few as 21 cells to as many as 13792 cells per cluster, were isolated by a graph-based clustering approach and labeled according to lineage-specific markers following batch correction (**Figure 1B**). The cell distribution from eight different kidney subjects was visualized by uniform manifold approximation and projection (UMAP) (**Figure 1C**). To define the character of each cell cluster, differential expression analysis was carried out to identify mutually exclusive sets of genes and therefore established markers of particular cell types. Enrichment of different cell clusters was calculated for each subject respectively (**Figure 1D**). The top 20 most differentially expressed markers in

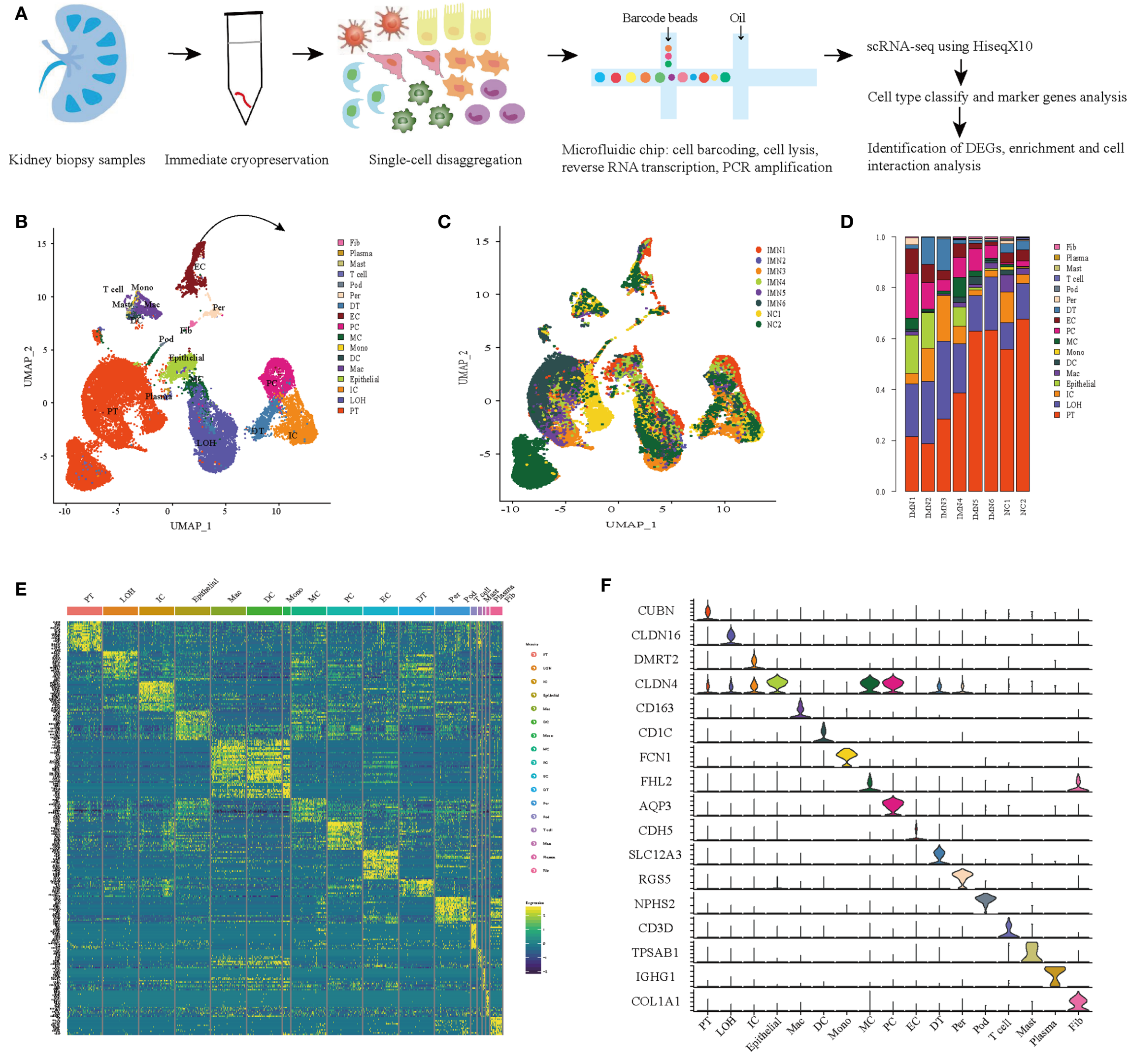


FIGURE 1 | Cell lineage analysis by comprehensive single-cell RNA-sequencing in anti-PLA2R positive IMN and control subjects. **(A)** Schematic of the scRNA-seq pipeline. Kidney samples from patients with IMN (n=6) or healthy control subjects (n=2) were collected at the time of clinically indicated renal biopsy or live kidney donation, respectively. Kidney biopsies were enzymatically disaggregated into single-cell suspensions and loaded onto a microfluidic device for cell barcoding, cell lysis, reverse RNA transcription, and then scRNA-seq as well as various other analyses. **(B)** Seventeen distinct cell clusters were visualized by UMAP plotting, with each cell color-coded for its associated subtypes. The color of the cells represents group origin. **(C)** UMAP plot of cell clusters from different subjects of IMN patients and control. The color of cells reflected the individual origin. **(D)** Bar plots of the percent contribution of cell clusters in kidneys from different subjects. Blocks represented different subjects, and block height was in proportion to the number of cells. **(E)** Heatmap of the top 20 most differentially expressed genes in each cluster to identify mutually exclusive gene sets, which were then used to determine the cell lineage of each cluster. Each column represented a cell cluster, and each row corresponded to a marker gene for the individual cluster. Transcript abundance ranges from low (purple) to high (yellow). **(F)** Violin plot of selected marker genes that identified the clusters generated by UMAP plotting. It was colored by different cell subtypes. PT, proximal tubule cells; LOH, loop of Henle cells; PC, principal cells; IC, intercalated cells; DT, distal tubule cells; EC, endothelial cells; Pod, podocytes; MC, mesangial cell; DC, dendritic cells; Mac, macrophages; Mono, monocytes; Fib, fibroblasts; Per, pericyte.

each cluster were shown in **Figure 1E**, and the selected cell lineage-specific marker gene was displayed in **Figure 1F**. For example, endothelial cells uniquely expressed *CDH5* and *KDR*, podocytes uniquely expressed *NPHS2* and *PODXL*, whereas mesangial cells expressed *FN1* and *FHL2*. Pericyte was labeled by *RGS5* and *ACTA2*. In addition, proximal tubule cells distinctly expressed *CUBN*, distal tubule cells distinctly expressed *SLC12A3*, whereas the loop of Henle cells uniquely expressed *CLDN16*. *DMRT2* and *SLC4A1* were defined as a cell-type specific marker for intercalated cells while *AQP3* were expressed specifically in principal cells. Fibroblasts expressed many genes encoding extracellular matrix proteins including *COL1A1* and *DCN*.

The comprehensive and detailed cell-lineage-specific marker genes of different kidney cells were displayed in **Table 1**.

Identification of DEGs and Enrichment Analysis in the Kidney Cells of Anti-PLA2R Positive IMN Subjects

To explore gene expression changes in kidney parenchymal cells, we performed differential expression analysis of transcriptomes between IMN patients and healthy donors. DEGs of kidney cells from the glomerulus and tubules were provided in **Datasets 1** and **2**, respectively. We defined representative DEGs in glomerular intrinsic cells (**Figure 2A**) as well as tubular intrinsic cells (**Figure 2B**) by comparing the transcriptional profile between IMN and control subjects.

Mesangial cells (MCs) of IMN highly expressed *IFI6* and *ATF3*. *IFI6*, a gene induced by interferon, which is associated with the regulation of apoptosis and type I interferon signaling pathway (24), was upregulated in MCs of IMN. Mesangial cells of IMN highly expressed ATF3, which was demonstrated to promote sublytic C5b-9-induced MCs apoptosis through up-regulation of *GADD45A* and *KLF6* gene expression (25). Besides, *KLF4*, *UBB, UBC*, and *DUSP1* were upregulated in endothelial cells of IMN. Endothelial *KLF4* mediated the protective effect of statins through regulating the expression of cell adhesion molecules and concomitant recruitment of inflammatory cells (26), which was implied by GO analysis in our study. *UBB* and *UBC*, both the important component of the ubiquitin pathway, were elevated in endothelial cells of IMN. They may play an essential role in the regulation of cell cycle, signal transduction as well as programmed cell death (27). Endothelial cells expressed *DUSP1*, a kind of two-way specific threonine/tyrosine phosphatase that regulates the mitogen-activated protein kinase (MAPK) signaling pathway by dephosphorylation of threonine/serine and tyrosine residues on its target (28) and regulates cell proliferation and cell growth cycle. Pericytes, a multifunctional cell-type of the kidney, highly expressed *FOS* that might play a role in the proliferation of pericytes and respond to PDGF-BB stimulation by phosphorylating both the PDGF receptor and the MAP kinase ERK-1/2 (29). Pericytes expressed CCL2 which might play a role in pericyte activation, proliferation, and differentiation into myofibroblasts during progressive kidney injury (30). GO enrichment analysis showed that DEGs were enriched in the regulation of apoptosis and type I interferon signaling pathway in mesangial cells, the regulation of programmed cell death, and various cytokine-mediated signaling pathways in endothelial cells and the regulation of protein modification in pericytes (**Figure 2D**), while KEGG enrichment analysis revealed that DEGs were mainly associated with the IL-17 signaling, TNF signaling, NOD-like receptor signaling and MAPK signaling pathway in endothelial cells as well as pericytes (**Figure 2E**).

DEGs upregulated in proximal tubules cells (PT) such a*s* *HSPA1A*, *TNFAIP3*, *KNG1*, and *TMSB4X*, were enriched in the regulation of cell proliferation, adhesion, programmed cell death, and response to cytokines in GO enrichment analysis (**Figure 2D**), whereas *NFKBIA*, *CXCL2*, *JUN*, *BIRC3* DEGs were enriched by KEGG enrichment analysis and participate in IL-17 signaling, TNF signaling, NOD-like receptor signaling, and NF-kappa B signaling (**Figure 2E**). Comparison of the DEGs in the distal tubule cells (DT) displayed enrichment of genes involved in oxidative phosphorylation, ATP metabolic process, and cation transmembrane transport (**Figure 2D**). The loop of Henle cells (LOH) of IMN had increased expression of *PLAU*, *KNG1*, *EEF2*, and *CAT*, which contribute to neutrophil-mediated immunity, exocytosis, and adherens junction (**Figure 2D**). A total of 2160 and 2194 cells were present in the principal cells (PC) and intercalated cells (IC), respectively. As illustrated in **Figures 2D, E**, DEGs of PCs between IMN and control subjects were enriched in IL-17 signaling, TNF signaling, NOD-like receptor signaling, and pattern recognition receptor signaling, of which *NFKBIA*, *JUN*, *CCL2*, *UBC* were involved. In addition to the genes responsible for acid secretion, including *SLC4A1*, *CLCNKB*, and *ATP6V1F*, DEGs of ICs were also enriched involving in adherens junction, oxidative phosphorylation, and organic compound catabolic process (**Figures 2D, E**).

Six clusters of leukocytes in the kidney of IMN were identified according to cell-specific differential genes, which were composed of dendritic cells (DC), macrophages, monocytes, mast cells, plasma cells, and T cells. The DEGs dataset of these leukocytes is displayed in **Figure 2C** and **Dataset 3**, except for monocytes, which were unable to be analyzed due to insufficient cell numbers. Also, the deficiency of meaningful differential genes in plasma cells, T cells, and mast cells was probably owing to the technical limits of isolating insufficient corresponding cells, since previous studies have confirmed their contributions in IMN (31–33). Macrophages highly expressed genes responsible for the regulation of leukocyte activation (*ZFP36L2*, *AXL*, and *RPS3*), inflammatory response (*KNG1*, *ELF3*, *CXCR4*, and *LY86*), and the antigen processing and presentation as well as immune response-regulating signaling pathway (*HLA-DPA1*, *HLA-DPB1*, and *HLA-DRB1*) (**Figures 2D, E**). Detailed information on the expression of the genes discussed above was provided in **Dataset 4**.

TABLE 1 | Cell-lineage-specific marker genes of different kidney cells.

Cell Type	Abbreviation	Marker genes
Proximal tubule cells	PT	CUBN, SLC13A1, LRP2, ALDOB
Mesangial cells	MC	FHL2, FN1, MYL9, CTGF
Podocytes	Pod	NPHS2, PODXL, PTPRO
Loop of Henle cells	LOH	UMOD, SLC12A1, CLDN16
Distal tubule cells	DT	CALB1, SLC12A3
Intercalated cells	IC	SLC4A1, ATP6V0D2, FOXI1, DMRT2
Principal cells	PC	AQP2, AQP3, GATA3
Epithelial cells	Epi	EPCAM, KRT8, CLDN4
Endothelial cells	EC	CDH5, PECAM1, KDR, CLDN5
Fibroblasts	Fib	COL1A1, DCN, LUM
Pericytes	Per	RGS5, ACTA2, MCAM, PDGFRB
T cells	T cell	CD3D, TRBC1, CD3E
Plasma cells	Plasma	IGHG1, JCHAIN, MZB1
Mast cells	Mast	TPSAB1, TPSB2, CPA3
Macrophages	Mac	MRC1, CD68, CD163, C1QA, IL1B
Monocytes	Mono	LYZ, CD14, VCAN, FCN1
Dendritic cells	DC	CD1C, FCER1A, CLEC10A, IRF8

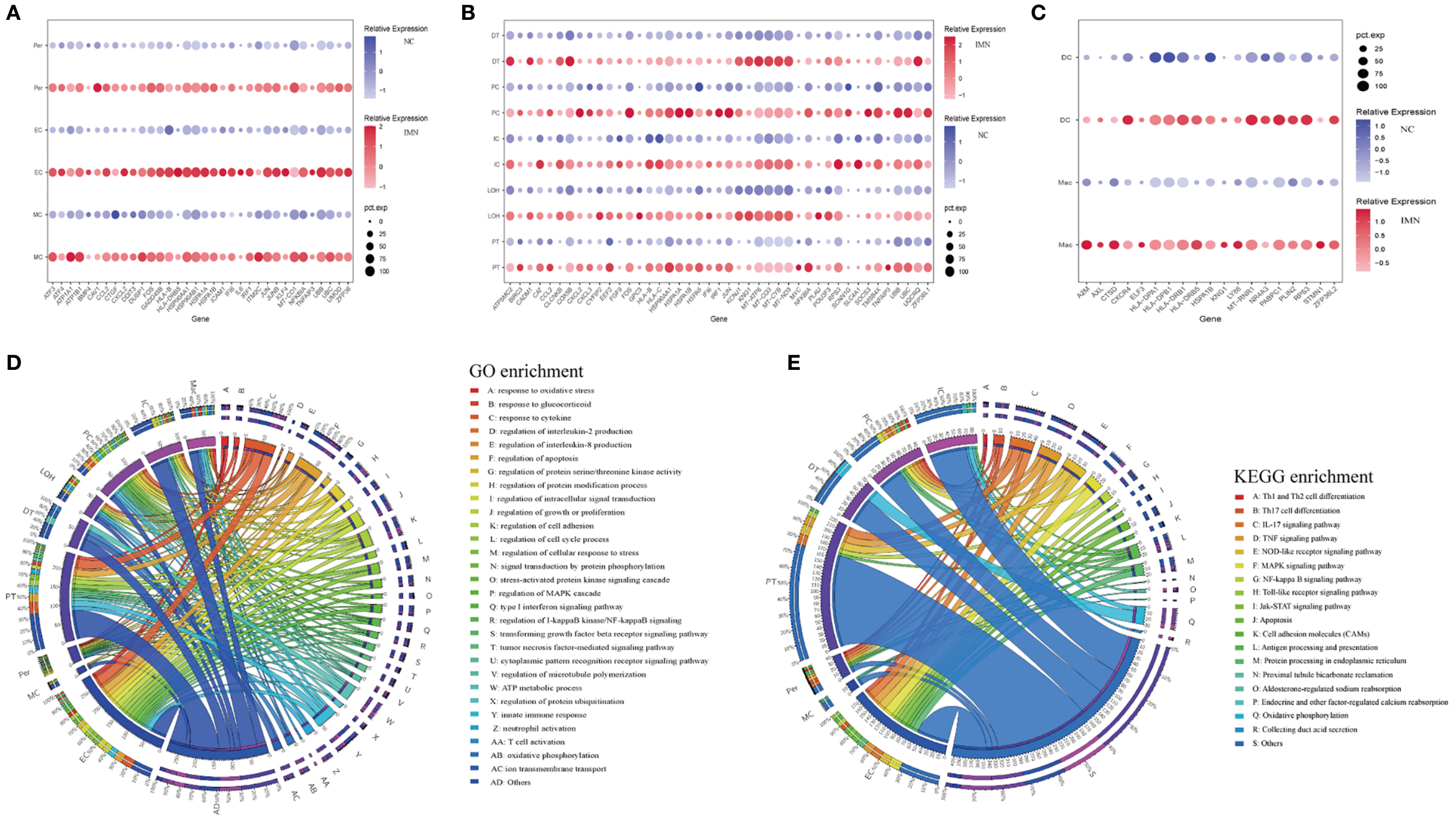


FIGURE 2 | DEGs and enrichment analysis in the kidney cells of anti-PLA2R positive IMN and control subjects. **(A)** Representative DEGs in mesangial cells, endothelial cells, and pericytes comparing the IMN patients to healthy donor control. pct.exp: percentage of cells expressing gene. **(B)** Representative DEGs in proximal tubule cells, distal tubule cells, loop of Henle cells, principal cells, and intercalated cells between IMN patients and control. **(C)** Representative DEGs in immune cells between IMN patients and control. **(D, E)** GO and KEGG enrichment shows the biological processes or signal pathways involved in different kidney cells, respectively. The left side of the circle represents different cell types, while the right side represents different biological processes or signaling pathways. The inner-circle represents gene numbers involved in cells or biological processes and signaling pathways, whereas the outer-circle represents the proportion of each cell type in biological processes and signaling pathways or the proportion of biological processes and signaling pathways in kidney cells.

Cell-Cell Crosstalk in Anti-PLA2R Positive IMN Through Ligand-Receptor Interactions

To explore the interactions and signaling network of different cell subsets in IMN, we performed ligand-receptor analysis. **Figure 3A** displayed the potential interactions of receptors and ligands in different cell types of kidneys. *CXCL1*, *CXCL8*, or *CCL2* expressed by mesangial cells interacted with *ACKR1* in endothelial cells (**Figure 3B**), which may participate in neutrophil/macrophage infiltration and inflammation response (34). FGF1 expressed by podocytes might ameliorate chronic kidney disease *via* PI3K/AKT mediated suppression of oxidative stress and inflammation (35) under the expression of FGFR1 in mesangial cells (**Figure 3C**). Moreover, PT expressed PTPRK, an important cell-cell adhesion regulator realized by reversible phosphorylation of protein tyrosine residues (36). We found it may interact with BMP7 from mesangial cells (**Figure 3D**). EGF, expressed by the loop of Henle cells, interacts with EGFR or

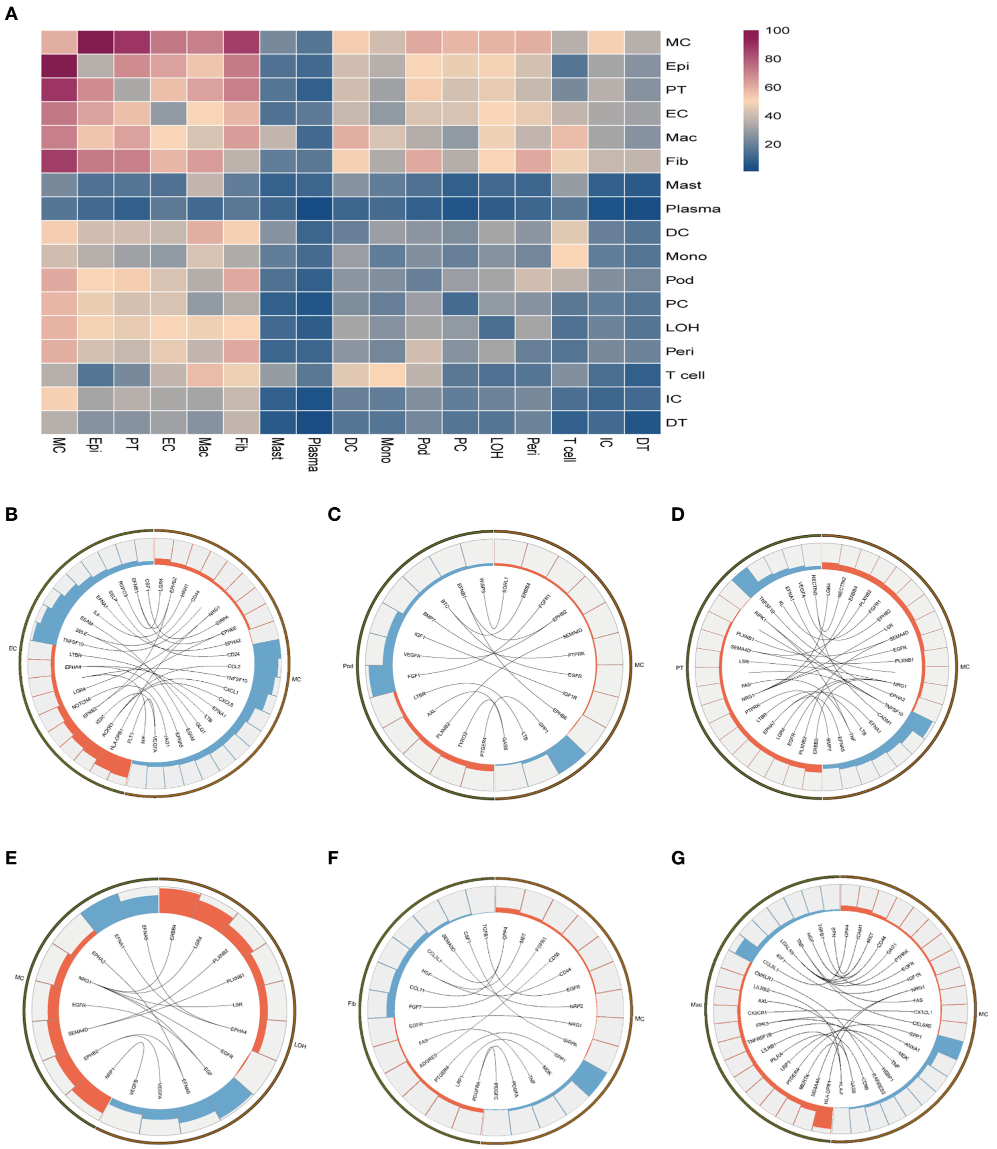


FIGURE 3 | Possible ligand-receptor interactions between different cell types in the kidney of anti-PLA2R positive IMN patients. **(A)** Ligand-receptor signaling pathways between cell clusters in the kidney. Cell-cell crosstalk frequency ranges from low (blue) to high (purple). **(B)** Representative ligand-receptor interactions between mesangial cells and endothelial cells. **(C)** Representative ligand-receptor interactions between mesangial cells and podocytes. **(D)** Representative ligand-receptor interactions between mesangial cells and proximal tubule cells. **(E)** Representative ligand-receptor interactions between mesangial cells and loop of Henle cells. **(F)** Representative ligand-receptor interactions between mesangial cells and fibroblasts. **(G)** Representative ligand-receptor interactions between mesangial cells and macrophages. Lines represented interrelations between the mesangial cells and other cells. Lines between the ligand and conjunct receptors were shown. Only IMN patients (n=6) were analyzed.

NGR1 in mesangial cells (**Figure 3E**), which probably plays a role in cell proliferation (37). Mesangial cells highly expressed SPP1 and become interaction pairs with PTGER4 from fibroblasts (**Figure 3F**), possibly involved in the activation of T cells. In addition to PTGER4 in fibroblasts, we found SPP1 in mesangial cells might interact with PTGER4 in macrophages (**Figure 3G**). The remaining information of cell-cell crosstalk was all displayed in **Figure 4**.

DEGs and Enrichment Analysis From Anti-PLA2R Positive IMN Patients Between Massive Proteinuria Group and Non-Massive Proteinuria Group

We displayed cell distribution between the massive proteinuria group and the non-massive proteinuria group from IMN patients by a graph-based clustering approach (**Figure 4A**). The grouping was based on whether proteinuria reached the scope of nephrotic syndrome, which is 3.5g per 24h. We then compared DEGs of proteinuria between the massive proteinuria group and the non-massive proteinuria group from IMN patients. The more specific information of DEGs from various kidney cells in two groups of IMN was shown in **Dataset 5**. ECs, MCs, PTs, LOHs, DTs, as well as ICs in IMN patients with massive proteinuria had elevated expression of *KLF6* that might participate in the glomerular mesangial cell proliferation, ECM accumulation, and proteinuria secretion (38). All the intrinsic kidney cells except for podocytes, pericytes, fibroblasts, and epithelial cells in IMN patients with massive proteinuria highly expressed SOCS3, a cytokine-inducible protein that might contribute to the regulation of receptor signaling in immune complex glomerulonephritis, in parallel with proteinuria and kidney lesions (39). Besides, all tubular cells in IMN patients with massive proteinuria had a significant expression of MMP7, which was secreted as a soluble protein from the tubules to the glomeruli and mediated the impairment of slit diaphragm integrity, leading to podocyte dysfunction and increased proteinuria (40). This suggests MMP-7 might be the key mediator of tubular-to-glomerular crosstalk that promotes proteinuria and CKD progression.

As exhibited in **Figures 4B, C**, DEGs between the massive proteinuria group and the non-massive proteinuria group from IMN patients were enriched in several common biological processes. For example, the regulation of apoptosis was enriched in MCs, LOHs, PTs, DTs, PCs, and ICs, whereas adherens junctions were enriched in PTs, LOHs, and DTs. Besides, the overexpressed genes in MCs, PTs, DTs, PCs, and ICs were mainly enriched in the response to cytokines, while most of the kidney parenchymal cells were enriched in the regulation of inflammation and immune response including TNF signaling, NOD-like receptor signaling, MAPK signaling as well as IL-17 signaling pathways.

DISCUSSION

Our understanding of the pathogenesis of IMN is limited by an incomplete molecular characterization of the cell types in the kidney and interaction between the cells. Given the organ-specific immunological characteristics of IMN, we performed unbiased single-cell RNA sequencing for the first time and

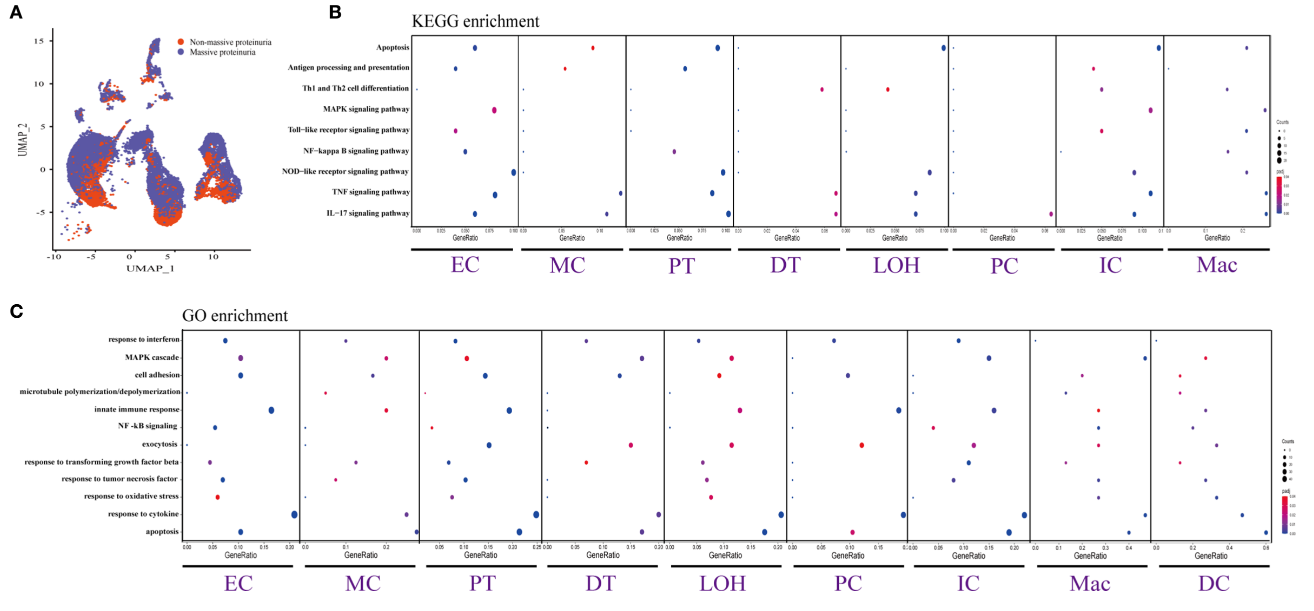


FIGURE 4 | DEGs involved in biological process and intercellular signaling in the kidney from anti-PLA2R positive IMN patients with massive proteinuria compared to non-massive proteinuria. **(A)** UMAP plot of cell clusters between massive proteinuria and non-massive proteinuria. The color of cells reflected different groups. **(B, C)** KEGG and GO enrichment analysis showed that upregulated DEGs were mostly involved in apoptosis, cell adhesion, and regulation of inflammation and the immune response in massive proteinuria patients, compared with non-massive proteinuria subjects.

identified all previously described cell types in the kidney. The DEGs in most kidney parenchymal cells were primarily enriched in the regulation of inflammation and immune response including IL-17 signaling, TNF signaling, NOD-like receptor signaling as well as MAPK signaling. Besides, the cell-cell crosstalk highlighted the extensive communication of mesangial cells, which infers great importance in IMN.

We provided abundant information of cell-type-specific gene expression and distinct signaling pathways by analysis of DEGs as well as enrichment. Glomerular cells including MCs, ECs, and pericytes, primarily participated in the regulation of programmed cell death, inflammatory process, and immune regulation, whereas tubular cells are mainly involved in adherens junction, oxidative phosphorylation as well as regulation of inflammation and immunity. *HSPA1A*, *ATF3*, *IFI6*, and *ITM2C*, which were significantly expressed in all glomerular cells and enriched in regulation of apoptosis, have not yet been implicated in IMN pathogenesis. Also, DEGs of the ECs, pericytes, PTs, and PCs, are related to the inflammatory process and immunity. Particularly, Th17 cells are key players in kidney autoimmunity by mediating fundamental inflammatory cascades and thereby may be of vital importance in IMN (41, 42). Previously, studies reported that IL-17 has several direct effects on kidney parenchymal cells facilitating leukocyte transmigration, promoting interaction with T-cells, and impacting kidney integrity (43, 44). These effects are inherent to the pathophysiological cascade in kidney autoimmunity (45). For example, a study demonstrated that tubular epithelial cells showed signs of disrupted cell-cell junctional integrity and loss of E-cadherin expression after exposure to IL-17 (44, 46), which is consistent with our scRNA-seq findings in kidney interstitial cells. We also observed extensive enrichment of NOD-like receptors (NLRs) signaling in glomerular as well as tubular cells. NLRs are recently identified intracellular PRRs that are essential to innate immune responses and tissue homeostasis. Emerging evidence suggested a potential role of NLRs in kidney disease (47, 48). Expression of Nod1, Nod2, or RICK induces NF-κB activation (49). In addition to NF-κB, Nod1 and Nod2 mediate the activation of JNK and p38 in response to microbial ligands (50, 51), which are expected to participate in the transcriptional activation of proinflammatory genes. However, to the best of our knowledge, none of the genes or pathways discussed above were explored in IMN patients. Thus, further studies are needed to validate our results and provide novel insights into the pathogenesis of human IMN.

The cell-cell crosstalk through ligand-receptor interactions was reported to show considerable importance in anti-PLA2R positive IMN pathogenesis (52). We focused on the regulation of inflammation and immunity between different cells, especially mesangial cells for their extensive communication with other cells in our study. A study demonstrated that lysophosphatidylcholine might stimulate EGF receptor transactivation and downstream MAP kinase signaling resulting in mesangial hypercellularity (53), while EGF similarly stimulated MAPK (ERK1/2) in HK-2 cells and consequently mediate cell proliferation (37), which might imply crosstalk between glomerular and tubular cells. Macrophages highly expressed *PTGER4*, which has been shown to drive the differentiation of Th1 cells and proliferation of Th17 cells (54). While SPP1, also known as osteopontin (OPN*)*, was upregulated in mesangial cells. Coincidentally, a study demonstrated that OPN functionally activates DCs and induces their differentiation toward a Th1-polarizing phenotype (55), which implies the possibility of communications between macrophages and mesangial cells. Furthermore, there are some hints that *CXCL1*, *CXCL8*, or *CCL2* expressed by mesangial cells interacted with *ACKR1* in endothelial cells. ACKR1, better known as Duffy antigen receptor for chemokines (DARC or Duffy), is usually thought to regulate the innate and adaptive immune response by acting as a chemokine reservoir or scavenger, and it seems to be a negative regulator of inflammation and immunological stimuli through combination with chemokines including CXCL1, CXCL8 and CCL2 (56). Chaudhuri et al. also demonstrated that vascular endothelial cells may induce Duffy protein to regulate leukocytes and chemokine trafficking (57). The communication of cells in the kidney is highly dynamic but more efforts are essential to elucidate the precisely controlled process.

As a crucial clinical indicator in various kidney diseases, proteinuria drew our attention as a matter of course. The enrichment analysis revealed that most of the intrinsic kidney cells were involved in inflammatory pathways in IMN patients with massive proteinuria, suggesting substantial functions in the disease setting. MAPK signaling was associated with proteinuria. A recent study reported that p38 MAPK mediated secretory phospholipase A2 group IB-induced autophagy in podocytes and promoted podocyte injury *via* activation of the mTOR/ULK1^{ser757} signaling pathway, which consequently lead to proteinuria (58). ERK and p38 pathways also mediated activation of calcium-independent phospholipase A2γ, which plays an important role in complement C5b-9-induced glomerular epithelial cell (GEC) injury and proteinuria (59). TNF signaling was also involved in IMN progression with proteinuria. Active membranous glomerulonephritis leads to not only proteinuria but also increased urinary TNF excretion (60). However, inhibition of TNF signaling might attenuate kidney immune cell infiltration in experimental membranous nephropathy (61). Also, circulating tumor necrosis factor receptors (cTNFRs) were suggested to predict renal progression in patients with IMN accompanied by nephrotic syndrome (62).

However, there were several limitations to our study. First, the number of patients in this study was small, which may inevitably lead to individual differences. To reduce this difference, additional specimens are needed to verify the results of our study. Second, as the podocytes are particularly important and even the core for membranous nephropathy (63–65), podocytes were rarely detected in this study is one major limitation. Currently, it is still a major challenge for the capture of rare cells, such as podocytes in the application of single-cell transcriptomics technologies to kidney diseases. The podocytes have not been successfully annotated in several previous single-cell sequencing studies focusing on kidney diseases because of the very small proportion of podocytes after digestion of needle kidney biopsy tissue (66–69). In our study, we clearly captured and annotated the population of podocytes. There were 27 and 23 podocytes captured in 6 patients with IMN and 2 healthy donors, respectively. However, we only found one differentially expressed gene due to the small number of podocytes captured in the disease group as well as healthy control group. A further increase in throughput of the next generation of single-cell sequencing techniques or extracting the glomerular from kidney tissue before dissociation into single cells may prove efficient to capture enough podocytes for subsequent analysis (70). Third, as no other kidney disease control group was included in this study, the changes of DEGs are not specific for anti-PLA2R positive IMN and the changes may be generic to being proteinuria or glomerular inflammation.

Overall, scRNA-seq served as a feasible and valuable technique performed in IMN patients. We demonstrated cell-specific transcriptional profiles in the kidney, anti-PLA2R positive IMN-associated novel genes, signaling pathways involved, and potential pathogenesis concerning ligand-receptor interactions. A better understanding of the molecular mechanism in IMN will provide yet unexplored opportunities to develop new therapies for kidney diseases. The results discovered in our study will be further validated using tissue staining, functional studies *in vitro* using cell lines or primary human cells, and animal models of IMN.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: https://www.ncbi.nlm.nih.gov/, GSE171458.

AUTHOR CONTRIBUTIONS

YZ, QZ, CS, JO, and PE conceived and supervised the project. JX and WL performed all biopsy dissociations and single-cell experiments. XD, PJ, TM, WN, XA, GX, and YT assisted with patient consent and sample acquisition of IMN biopsies and assisted with patient consent and sample acquisition of live kidney donor tissue. Renal biopsy histology was evaluated by WL. Analysis was performed by YZ, JX, XA, and RT. JX, YZ, and QZ prepared and wrote the manuscript. JO and PE revised the paper. All authors contributed to the article and approved the submitted version.

CONFERENCE PRESENTATION

Parts of the present study have been accepted as a Mini-Oral at the 58th ERA-EDTA Congress, which will be organized from June 5 to 8, 2021.

FUNDING

This work was funded by the National Key R&D Program of China (2020YFC2005000 to XX), the National Natural Science Foundation of China (81800641 to TM), the Key Research and Development Program of Hunan province (2018WK2060 to XX and 2020WK2008 to YZ), the science and technology innovation Program of Hunan Province (2020RC5002 to JO), the Natural Science Foundation of Hunan Province (2020JJ6109 to CS and 2019JJ40515 to WN), Chinese Society of Nephrology (18020010780 to YZ).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.683330/full#supplementary-material

REFERENCES

1. Ayalon R, Beck LHJr. Membranous Nephropathy: Not Just a Disease for Adults. *Pediatr Nephrol* (2015) 30(1):31–9. doi: 10.1007/s00467-013-2717-z
2. Hou JH, Zhu HX, Zhou ML, Le WB, Zeng CH, Liang SS, et al. Changes in the Spectrum of Kidney Diseases: An Analysis of 40,759 Biopsy-Proven Cases From 2003 to 2014 in China. *Kidney Dis (Basel)* (2018) 4(1):10–9. doi: 10.1159/000484717
3. Haas M, Meehan SM, Karrison TG, Spargo BH. Changing Etiologies of Unexplained Adult Nephrotic Syndrome: A Comparison of Renal Biopsy Findings From 1976-1979 and 1995-1997. *Am J Kidney Dis* (1997) 30(5):621–31. doi: 10.1016/s0272-6386(97)90485-6
4. Hanko JB, Mullan RN, O'Rourke DM, McNamee PT, Maxwell AP, Courtney AE. The Changing Pattern of Adult Primary Glomerular Disease. *Nephrol Dial Transplant* (2009) 24(10):3050–4. doi: 10.1093/ndt/gfp254
5. Simon P, Ramee MP, Autuly V, Laruelle E, Charasse C, Cam G, et al. Epidemiology of Primary Glomerular Diseases in a French Region. Variations According to Period and Age. *Kidney Int* (1994) 46(4):1192–8. doi: 10.1038/ki.1994.384
6. Schieppati A, Mosconi L, Perna A, Mecca G, Bertani T, Garattini S, et al. Prognosis of Untreated Patients With Idiopathic Membranous Nephropathy. *N Engl J Med* (1993) 329(2):85–9. doi: 10.1056/NEJM199307083290203
7. Polanco N, Gutierrez E, Covarsi A, Ariza F, Carreno A, Vigil A, et al. Spontaneous Remission of Nephrotic Syndrome in Idiopathic Membranous Nephropathy. *J Am Soc Nephrol* (2010) 21(4):697–704. doi: 10.1681/ASN.2009080861
8. Polanco N, Gutierrez E, Rivera F, Castellanos I, Baltar J, Lorenzo D, et al. Spontaneous Remission of Nephrotic Syndrome in Membranous Nephropathy With Chronic Renal Impairment. *Nephrol Dial Transplant* (2012) 27(1):231–4. doi: 10.1093/ndt/gfr285
9. Ronco P, Debiec H. Molecular Pathogenesis of Membranous Nephropathy. *Annu Rev Pathol* (2020) 15:287–313. doi: 10.1146/annurev-pathol-020117-043811

10. Liu W, Gao C, Dai H, Zheng Y, Dong Z, Gao Y, et al. Immunological Pathogenesis of Membranous Nephropathy: Focus on PLA2R1 and Its Role. *Front Immunol* (2019) 10:1809. doi: 10.3389/fimmu.2019.01809
11. Francis JM, Beck LH Jr, Salant DJ. Membranous Nephropathy: A Journey From Bench to Bedside. *Am J Kidney Dis* (2016) 68(1):138–47. doi: 10.1053/j.ajkd.2016.01.030
12. Akiyama S, Akiyama M, Imai E, Ozaki T, Matsuo S, Maruyama S. Prevalence of Anti-Phospholipase A2 Receptor Antibodies in Japanese Patients With Membranous Nephropathy. *Clin Exp Nephrol* (2015) 19(4):653–60. doi: 10.1007/s10157-014-1054-2
13. Stanescu HC, Arcos-Burgos M, Medlar A, Bockenhauer D, Kottgen A, Dragomirescu L, et al. Risk HLA-DQA1 and PLA(2)R1 Alleles in Idiopathic Membranous Nephropathy. *N Engl J Med* (2011) 364(7):616–26. doi: 10.1056/NEJMoa1009742
14. Le WB, Shi JS, Zhang T, Liu L, Qin HZ, Liang S, et al. Hla-DRB1*15:01 and HLA-DRB3*02:02 in PLA2R-Related Membranous Nephropathy. *J Am Soc Nephrol* (2017) 28(5):1642–50. doi: 10.1681/ASN.2016060644
15. Debiec H, Ronco P. Immunopathogenesis of Membranous Nephropathy: An Update. *Semin Immunopathol* (2014) 36(4):381–97. doi: 10.1007/s00281-014-0423-y
16. Ma H, Sandor DG, Beck LHJr. The Role of Complement in Membranous Nephropathy. *Semin Nephrol* (2013) 33(6):531–42. doi: 10.1016/j.semnephrol.2013.08.004
17. Saliba AE, Westermann AJ, Gorski SA, Vogel J. Single-Cell RNA-Seq: Advances and Future Challenges. *Nucleic Acids Res* (2014) 42(14):8845–60. doi: 10.1093/nar/gku555
18. Kiryluk K, Bomback AS, Cheng YL, Xu K, Camara PG, Rabadan R, et al. Precision Medicine for Acute Kidney Injury (AKI): Redefining AKI by Agnostic Kidney Tissue Interrogation and Genetics. *Semin Nephrol* (2018) 38(1):40–51. doi: 10.1016/j.semnephrol.2017.09.006
19. Arazi A, Rao DA, Berthier CC, Davidson A, Liu Y, Hoover PJ, et al. The Immune Cell Landscape in Kidneys of Patients With Lupus Nephritis. *Nat Immunol* (2019) 20(7):902–14. doi: 10.1038/s41590-019-0398-x
20. Wilson PC, Wu H, Kirita Y, Uchimura K, Ledru N, Rennke HG, et al. The Single-Cell Transcriptomic Landscape of Early Human Diabetic Nephropathy. *Proc Natl Acad Sci USA* (2019) 116(39):19619–25. doi: 10.1073/pnas.1908706116
21. Hu J, Chen Z, Bao L, Zhou L, Hou Y, Liu L, et al. Single-Cell Transcriptome Analysis Reveals Intratumoral Heterogeneity in ccRCC, Which Results in Different Clinical Outcomes. *Mol Ther* (2020) 28(7):1658–72. doi: 10.1016/j.ymthe.2020.04.023
22. Bobart SA, De Vriese AS, Pawar AS, Zand L, Sethi S, Giesen C, et al. Noninvasive Diagnosis of Primary Membranous Nephropathy Using Phospholipase A2 Receptor Antibodies. *Kidney Int* (2019) 95(2):429–38. doi: 10.1016/j.kint.2018.10.021
23. Markowitz GS. Membranous Glomerulopathy: Emphasis on Secondary Forms and Disease Variants. *Adv Anat Pathol* (2001) 8(3):119–25. doi: 10.1097/00125480-200105000-00001
24. Schoggins JW, Wilson SJ, Panis M, Murphy MY, Jones CT, Bieniasz P, et al. A Diverse Range of Gene Products are Effectors of the Type I Interferon Antiviral Response. *Nature* (2011) 472(7344):481–5. doi: 10.1038/nature09907
25. Xu K, Zhou Y, Qiu W, Liu X, Xia M, Liu L, et al. Activating Transcription Factor 3 (ATF3) Promotes Sublytic C5b-9-Induced Glomerular Mesangial Cells Apoptosis Through Up-Regulation of Gadd45alpha and KLF6 Gene Expression. *Immunobiology* (2011) 216(8):871–81. doi: 10.1016/j.imbio.2011.02.005
26. Yoshida T, Yamashita M, Iwai M, Hayashi M. Endothelial Kruppel-Like Factor 4 Mediates the Protective Effect of Statins Against Ischemic AKI. *J Am Soc Nephrol* (2016) 27(5):1379–88. doi: 10.1681/ASN.2015040460
27. Popovic D, Vucic D, Dikic I. Ubiquitination in Disease Pathogenesis and Treatment. *Nat Med* (2014) 20(11):1242–53. doi: 10.1038/nm.3739
28. Bermudez O, Pages G, Gimond C. The Dual-Specificity MAP Kinase Phosphatases: Critical Roles in Development and Cancer. *Am J Physiol Cell Physiol* (2010) 299(2):C189–202. doi: 10.1152/ajpcell.00347.2009
29. Tomkowicz B, Rybinski K, Sebeck D, Sass P, Nicolaides NC, Grasso L, et al. Endosialin/TEM-1/CD248 Regulates Pericyte Proliferation Through PDGF Receptor Signaling. *Cancer Biol Ther* (2010) 9(11):908–15. doi: 10.4161/cbt.9.11.11731
30. Chen YT, Chang FC, Wu CF, Chou YH, Hsu HL, Chiang WC, et al. Platelet-Derived Growth Factor Receptor Signaling Activates Pericyte-Myofibroblast Transition in Obstructive and Post-Ischemic Kidney Fibrosis. *Kidney Int* (2011) 80(11):1170–81. doi: 10.1038/ki.2011.208
31. Zhang Y, Jin Y, Guan Z, Li H, Su Z, Xie C, et al. The Landscape and Prognosis Potential of the T-Cell Repertoire in Membranous Nephropathy. *Front Immunol* (2020) 11:387. doi: 10.3389/fimmu.2020.00387
32. Rosenzwajg M, Languille E, Debiec H, Hygino J, Dahan K, Simon T, et al. B- and T-cell Subpopulations in Patients With Severe Idiopathic Membranous Nephropathy may Predict an Early Response to Rituximab. *Kidney Int* (2017) 92(1):227–37. doi: 10.1016/j.kint.2017.01.012
33. Tozzoli R. Receptor Autoimmunity: Diagnostic and Therapeutic Implications. *Auto Immun Highlights* (2020) 11(1):1. doi: 10.1186/s13317-019-0125-5
34. Chung AC, Lan HY. Chemokines in Renal Injury. *J Am Soc Nephrol* (2011) 22 (5):802–9. doi: 10.1681/ASN.2010050510
35. Wang D, Jin M, Zhao X, Zhao T, Lin W, He Z, et al. Fgf1(DeltaHBS) Ameliorates Chronic Kidney Disease Via PI3K/AKT Mediated Suppression of Oxidative Stress and Inflammation. *Cell Death Dis* (2019) 10(6):464. doi: 10.1038/s41419-019-1696-9
36. Fearnley GW, Young KA, Edgar JR, Antrobus R, Hay IM, Liang WC, et al. The Homophilic Receptor PTPRK Selectively Dephosphorylates Multiple Junctional Regulators to Promote Cell-Cell Adhesion. *Elife* (2019) 8:e44597. doi: 10.7554/eLife.44597
37. Zepeda-Orozco D, Wen HM, Hamilton BA, Raikwar NS, Thomas CP. EGF Regulation of Proximal Tubule Cell Proliferation and VEGF-A Secretion. *Physiol Rep* (2017) 5(18):e13453. doi: 10.14814/phy2.13453
38. Yu T, Gong Y, Liu Y, Xia L, Zhao C, Liu L, et al. Klf6 Acetylation Promotes Sublytic C5b-9-Induced Production of MCP-1 and RANTES in Experimental Mesangial Proliferative Glomerulonephritis. *Int J Biol Sci* (2020) 16(13):2340–56. doi: 10.7150/ijbs.46573
39. Gomez-Guerrero C, Lopez-Franco O, Sanjuan G, Hernandez-Vargas P, Suzuki Y, Ortiz-Munoz G, et al. Suppressors of Cytokine Signaling Regulate Fc Receptor Signaling and Cell Activation During Immune Renal Injury. *J Immunol* (2004) 172(11):6969–77. doi: 10.4049/jimmunol.172.11.6969
40. Tan RJ, Li Y, Rush BM, Cerqueira DM, Zhou D, Fu H, et al. Tubular Injury Triggers Podocyte Dysfunction by Beta-Catenin-Driven Release of MMP-7. *JCI Insight* (2019) 4(24):e122399. doi: 10.1172/jci.insight.122399
41. Dolff S, Witzke O, Wilde B. Th17 Cells in Renal Inflammation and Autoimmunity. *Autoimmun Rev* (2019) 18(2):129–36. doi: 10.1016/j.autrev.2018.08.006
42. Krebs CF, Panzer U. Plasticity and Heterogeneity of Th17 in Immune-Mediated Kidney Diseases. *J Autoimmun* (2018) 87:61–8. doi: 10.1016/j.jaut.2017.12.005
43. Kuo HL, Huang CC, Lin TY, Lin CY. Il-17 and CD40 Ligand Synergistically Stimulate the Chronicity of Diabetic Nephropathy. *Nephrol Dial Transplant* (2018) 33(2):248–56. doi: 10.1093/ndt/gfw397
44. Dudas PL, Sague SL, Elloso MM, Farrell FX. Proinflammatory/Profibrotic Effects of interleukin-17A on Human Proximal Tubule Epithelium. *Nephron Exp Nephrol* (2011) 117(4):e114–23. doi: 10.1159/000320177
45. Ghali JR, Holdsworth SR, Kitching AR. Targeting IL-17 and IL-23 in Immune Mediated Renal Disease. *Curr Med Chem* (2015) 22(38):4341–65. doi: 10.2174/0929867322666151030163022
46. Liu L, Li FG, Yang M, Wang L, Chen Y, Wang L, et al. Effect of Pro-Inflammatory interleukin-17A on Epithelial Cell Phenotype Inversion in HK-2 Cells In Vitro. *Eur Cytokine Netw* (2016) 27(2):27–33. doi: 10.1684/ecn.2016.0373
47. Leemans JC, Kors L, Anders HJ, Florquin S. Pattern Recognition Receptors and the Inflammasome in Kidney Disease. *Nat Rev Nephrol* (2014) 10(7):398–414. doi: 10.1038/nrneph.2014.91
48. Wang X, Yi F. The Nucleotide Oligomerization Domain-Like Receptors in Kidney Injury. *Kidney Dis (Basel)* (2016) 2(1):28–36. doi: 10.1159/000444736
49. Inohara N, Koseki T, del Peso L, Hu Y, Yee C, Chen S, et al. Nod1, an Apaf-1-like Activator of Caspase-9 and Nuclear Factor-Kappab. *J Biol Chem* (1999) 274(21):14560–7. doi: 10.1074/jbc.274.21.14560
50. Girardin SE, Tournebize R, Mavris M, Page AL, Li X, Stark GR, et al. CARD4/Nod1 Mediates NF-kappaB and JNK Activation by Invasive Shigella Flexneri. *EMBO Rep* (2001) 2(8):736–42. doi: 10.1093/embo-reports/kve155

51. Kobayashi K, Inohara N, Hernandez LD, Galan JE, Nunez G, Janeway CA, et al. Rick/Rip2/CARDIAK Mediates Signalling for Receptors of the Innate and Adaptive Immune Systems. *Nature* (2002) 416(6877):194–9. doi: 10.1038/416194a
52. Cassis P, Zoja C, Perico L, Remuzzi G. A Preclinical Overview of Emerging Therapeutic Targets for Glomerular Diseases. *Expert Opin Ther Targets* (2019) 23(7):593–606. doi: 10.1080/14728222.2019.1626827
53. Bassa BV, Noh JW, Ganji SH, Shin MK, Roh DD, Kamanna VS. Lysophosphatidylcholine Stimulates EGF Receptor Activation and Mesangial Cell Proliferation: Regulatory Role of SRC and PKC. *Biochim Biophys Acta* (2007) 1771(11):1364–71. doi: 10.1016/j.bbalip.2007.09.004
54. Yao C, Sakata D, Esaki Y, Li Y, Matsuoka T, Kuroiwa K, et al. Prostaglandin E2-EP4 Signaling Promotes Immune Inflammation Through Th1 Cell Differentiation and Th17 Cell Expansion. *Nat Med* (2009) 15(6):633–40. doi: 10.1038/nm.1968
55. Renkl AC, Wussler J, Ahrens T, Thoma K, Kon S, Uede T, et al. Osteopontin Functionally Activates Dendritic Cells and Induces Their Differentiation Toward a Th1-Polarizing Phenotype. *Blood* (2005) 106(3):946–55. doi: 10.1182/blood-2004-08-3228
56. Mantovani A, Bonecchi R, Locati M. Tuning Inflammation and Immunity by Chemokine Sequestration: Decoys and More. *Nat Rev Immunol* (2006) 6 (12):907–18. doi: 10.1038/nri1964
57. Chaudhuri A, Rodriguez M, Zbrzezna V, Luo H, Pogo AO, Banerjee D. Induction of Duffy Gene (FY) in Human Endothelial Cells and in Mouse. *Cytokine* (2003) 21(3):137–48. doi: 10.1016/s1043-4666(03)00033-4
58. Yang L, Wu Y, Lin S, Dai B, Chen H, Tao X, et al. sPLA2-IB and PLA2R Mediate Insufficient Autophagy and Contribute to Podocyte Injury in Idiopathic Membranous Nephropathy by Activation of the P38mapk/mTOR/ULK1(ser757) Signaling Pathway. *FASEB J* (2021) 35(2):e21170. doi: 10.1096/fj.202001143R
59. Elimam H, Papillon J, Takano T, Cybulsky AV. Complement-Mediated Activation of Calcium-Independent Phospholipase A2gamma: Role of Protein Kinases and Phosphorylation. *J Biol Chem* (2013) 288(6):3871–85. doi: 10.1074/jbc.M112.396614
60. Honkanen E, von Willebrand E, Teppo AM, Tornroth T, Gronhagen-Riska C. Adhesion Molecules and Urinary Tumor Necrosis Factor-Alpha in Idiopathic Membranous Glomerulonephritis. *Kidney Int* (1998) 53(4):909–17. doi: 10.1111/j.1523-1755.1998.00833.x
61. Huang YS, Fu SH, Lu KC, Chen JS, Hsieh HY, Sytwu HK, et al. Inhibition of Tumor Necrosis Factor Signaling Attenuates Renal Immune Cell Infiltration in Experimental Membranous Nephropathy. *Oncotarget* (2017) 8 (67):111631–41. doi: 10.18632/oncotarget.22881
62. Lee SM, Yang S, Cha RH, Kim M, An JN, Paik JH, et al. Circulating TNF Receptors are Significant Prognostic Biomarkers for Idiopathic Membranous Nephropathy. *PloS One* (2014) 9(8):e104354. doi: 10.1371/journal.pone.0104354
63. Ronco P, Debiec H. Pathophysiological Advances in Membranous Nephropathy: Time for a Shift in Patient's Care. *Lancet* (2015) 385 (9981):1983–92. doi: 10.1016/S0140-6736(15)60731-0
64. Alsharhan L, Beck LHJr. Membranous Nephropathy: Core Curriculum 2021. *Am J Kidney Dis* (2021) 77(3):440–53. doi: 10.1053/j.ajkd.2020.10.009
65. Xu Z, Chen L, Xiang H, Zhang C, Xiong J. Advances in Pathogenesis of Idiopathic Membranous Nephropathy. *Kidney Dis (Basel)* (2020) 6(5):330–45. doi: 10.1159/000507704
66. Wu H, Malone AF, Donnelly EL, Kirita Y, Uchimura K, Ramakrishnan SM, et al. Single-Cell Transcriptomics of a Human Kidney Allograft Biopsy Specimen Defines a Diverse Inflammatory Response. *J Am Soc Nephrol* (2018) 29(8):2069–80. doi: 10.1681/ASN.2018020125
67. Liu Y, Hu J, Liu D, Zhou S, Liao J, Liao G, et al. Single-Cell Analysis Reveals Immune Landscape in Kidneys of Patients With Chronic Transplant Rejection. *Theranostics* (2020) 10(19):8851–62. doi: 10.7150/thno.48201
68. Der E, Ranabothu S, Suryawanshi H, Akat KM, Clancy R, Morozov P, et al. Single Cell RNA Sequencing to Dissect the Molecular Heterogeneity in Lupus Nephritis. *JCI Insight* (2017) 2(9):e93009. doi: 10.1172/jci.insight.93009
69. Der E, Suryawanshi H, Morozov P, Kustagi M, Goilav B, Ranabothu S, et al. Tubular Cell and Keratinocyte Single-Cell Transcriptomics Applied to Lupus Nephritis Reveal Type I IFN and Fibrosis Relevant Pathways. *Nat Immunol* (2019) 20(7):915–27. doi: 10.1038/s41590-019-0386-1
70. Zheng Y, Lu P, Deng Y, Wen L, Wang Y, Ma X, et al. Single-Cell Transcriptomics Reveal Immune Mechanisms of the Onset and Progression of IgA Nephropathy. *Cell Rep* (2020) 33(12):108525. doi: 10.1016/j.celrep.2020.108525

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

REVIEW
published: 11 June 2021
doi: 10.3389/fimmu.2021.668045

Multi-Omics Approaches in Immunological Research

Xiaojing Chu[1,2,3†], Bowen Zhang[2,3†], Valerie A. C. M. Koeken[2,3,4], Manoj Kumar Gupta[2,3] and Yang Li[1,2,3,4*]*

[1] Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, Netherlands, [2] Department of Computational Biology for Individualised Medicine, Centre for Individualised Infection Medicine (CiiM), a joint venture between the Hannover Medical School and the Helmholtz Centre for Infection Research, Hannover, Germany, [3] TWINCORE, Centre for Experimental and Clinical Infection Research, a joint venture between the Hannover Medical School and the Helmholtz Centre for Infection Research, Hannover, Germany, [4] Department of Internal Medicine and Radboud Center for Infectious Diseases, Radboud University Medical Center, Nijmegen, Netherlands

OPEN ACCESS

Edited by:
Lam Cheung Tsoi,
University of Michigan, United States

Reviewed by:
Genhong Yao,
Nanjing Drum Tower Hospital, China
Leandro Balzano-Nogueira,
University of Florida, United States
Xianyong Yin,
University of Michigan, United States

****Correspondence:***
Xiaojing Chu
Tsuki1128@outlook.com
Yang Li
Yang.Li@helmholtz-hzi.de

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Autoimmune and Autoinflammatory Disorders, a section of the journal Frontiers in Immunology

Received: *15 February 2021*
Accepted: *28 May 2021*
Published: *11 June 2021*

Citation:
Chu X, Zhang B, Koeken VACM, Gupta MK and Li Y (2021) Multi-Omics Approaches in Immunological Research. Front. Immunol. 12:668045. doi: 10.3389/fimmu.2021.668045

The immune system plays a vital role in health and disease, and is regulated through a complex interactive network of many different immune cells and mediators. To understand the complexity of the immune system, we propose to apply a multi-omics approach in immunological research. This review provides a complete overview of available methodological approaches for the different omics data layers relevant for immunological research, including genetics, epigenetics, transcriptomics, proteomics, metabolomics, and cellomics. Thereafter, we describe the various methods for data analysis as well as how to integrate different layers of omics data. Finally, we discuss the possible applications of multi-omics studies and opportunities they provide for understanding the complex regulatory networks as well as immune variation in various immune-related diseases.

Keywords: multi-omics, systems immunology, integrative analysis, immune-related diseases, immune variation

INTRODUCTION

Infections cause millions of deaths each year, and the current COVID-19 pandemic underlines the devastating effects of these communicable diseases. At the same time, the incidence of immune-related diseases such as atherosclerosis (1) and autoimmune diseases such as type 1 diabetes mellitus (2) have been increasing. All these diseases are related to or mediated by the immune system. Thus, the immune system plays a vital role in health and disease, and it is our defense mechanism against harmful substances, infectious diseases and cancer. Within a properly functioning immune system, immune responses should be kept in a certain range, as both hypo-activation and hyper-activation lead to disorders of the immune system. Understanding how the immune system works and what causes the immune system disorders may help us to efficiently fight against immune-related diseases.

However, getting a comprehensive understanding of the immune system is a challenging task. First of all, the immune response is mediated through a complex interactive network of many different immune cells and molecules, such as cytokines, immunoglobulins, and metabolites. At the same time, this network is highly variable depending on the exact threat of the wide variety of pathogens and other substances it's responding to. To make things even more complex, the immune response to a certain stimulus or infection is highly variable between individuals, leading to population heterogeneity. This heterogeneity is exemplified by differences in severity of patients suffering from the same infectious disease (3), variability in vaccine efficacy (4), and variation in responses to the same medical treatment (5). Many factors contribute to the immune network and the inter-individual variation of immune responses, highlighting both the promise and the challenge of multi-omics studies.

Until now, omics data have been used in many immunological studies to identify the determinants of immune variation and molecular bases of the immune process in different population groups. Properly designed omics studies should make use of appropriate measurements as well as reasonable analytic approaches, which depend on their specific research question. Taking omics studies on COVID-19 as an example, a genome-wide association study revealed eight genetic regions to be associated with critical illness in COVID-19. By integrating both genome and transcriptome data, the authors prioritized one gene, *IFNAR2*, that might play a causal role in COVID-19 (6). Another study, focusing on transcriptome data of immune cells from the lung and blood, identified several pro-inflammatory immune pathways related to the pathogenesis of COVID-19 (7). A proteomics and metabolomics study investigated the changes in COVID-19 patient sera, and identified molecular changes implicating dysregulation in macrophage pathways, complement activation, and platelet degranulation, as well as suppression of metabolic pathways (8). A cellomics and single-cell transcriptome study also revealed dysregulation of the monocyte compartment as well as two neutrophils clusters specific to severe COVID-19 patients (9). Moreover, a study integrating single-cell transcriptome, cellomics, epigenome and proteome comprehensively characterized complex dynamic changes in immune cells. Their results disclose an elevation of IFN-activated megakaryocytes and erythroid cells, hypomethylations around immune signaling genes, and co-expression modules associated with clinical outcome (10). Additionally, a study on fecal fungal microbiota of COVID-19 patients showed enrichment of *Candia albicans* and a highly heterogeneous mycobiome configuration during hospitalization (11). From different angles, these studies make use of omics data to provide insights in the molecular pathology of COVID-19, which can eventually lead to improved therapeutic strategies.

In this review, we present an introduction to multi-omics studies to investigate immune function and variation. The review is split into three parts. In the first section, we describe in brief about the different layers of omics data relevant for immunological research, including genome, epigenome, transcriptome, proteome, metabolome, digestive system microbiota and cellomics (12) [also called cytomics (13)] (**Figure 1**), and the commonly used methodological approaches to measure these different types of omics data. We also discuss important considerations and recommendations for an appropriate study design. In the second section, we discuss how to analyze and integrate multiple omics platforms, including system genetic approaches to identify genetic factors, integration among multiple genetic profiles, as well as the integration and association with other omics data layers. We demonstrate how recent studies applied a multi-omics approach to the immune system researches, and we discuss the interpretation of results from different approaches and their importance in immunological studies. In the third section, we discuss the immunological subjects that need specific attention and may see progress in the next few years. As for detailed information on computational algorithms and models in multi-omics integration (14, 15), imputation on missing omics data (16), and strengths and limitations of system approaches in infectious diseases research (17), we refer readers to other recent reviews.

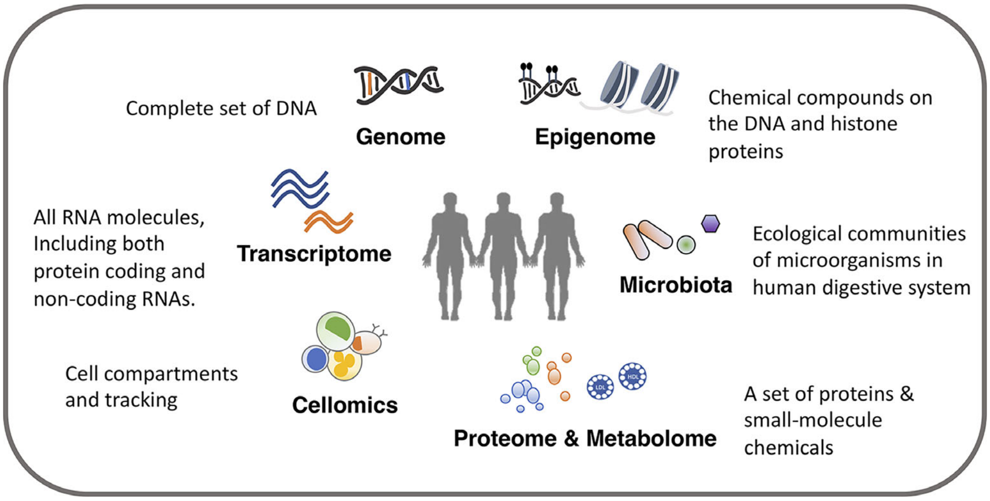

Complete set of DNA
Genome
Epigenome
Chemical compounds on the DNA and histone proteins
All RNA molecules, Including both protein coding and non-coding RNAs.
Transcriptome
Microbiota
Ecological communities of microorganisms in human digestive system
Cell compartments and tracking
Cellomics
Proteome & Metabolome
A set of proteins & small-molecule chemicals

FIGURE 1 | Overview of omics data.

MEASUREMENTS OF OMICS DATA

We can identify potential immunological mediators and study immune phenotypes with a wide range of omics comprising of various molecular and cellular phenotypes including genome, epigenome, transcriptome, proteome, metabolome, digestive system microbiota and cellular phenotypes such as cell composition (**Table 1**). A single omics data layer characterizes a specific biological process from one aspect, for example, transcriptome, but this can only provide insights on genes at a transcriptional level. To achieve a holistic picture of the immune system, a systematic collection of multi-omics data is often required. The tissue (or source) to be measured is another important aspect to be considered. For example, the genome is usually regarded as a stable feature for each individual and collected from an easily accessible tissue, such as blood. Only in some specific contexts, somatic mutations acquired after birth have to be considered and measured in specific tissues (18). However, many other types of omics, such as transcriptome, proteome and metabolome, vary between cell types and tissues. Therefore, it is important to consider the tissue in your experimental design and aim to get as close to the relevant tissue as possible.

Given the complexity of the immune system, there is no golden standard for what to collect in multi-omics studies. The necessary data depends on the research question and subjects. Understanding the different layers of omics data is helpful for setting up an appropriate study design. Therefore, in this part, we introduce features and categories of different omics, and describe important considerations when generating these data.

GENOME VARIATION MEASUREMENT

Genotyping detects diversity in the genome. It describes small variations, such as single-nucleotide polymorphisms (SNPs), insertion/deletions (InDels) as well as large-scale mutations such as insertions, deletions and amplifications. Genetic diversity can lead to variation in individual immune function (19).

To date, many techniques can be used for detecting genotypes, including DNA sequencing, DNA microarrays (also known as genotyping chips) and PCR-based methods. These approaches can be categorized based on their measurement scales (high-throughput *vs.* low-throughput methods) or based on whether they include unknown variants (discovery *vs.* screening methods). Classical sequencing-based approaches detect genetic variants in a nearly unbiased manner on the genome (whole-genome sequencing) or within the exome regions (whole-exome sequencing), including known or novel SNPs as well as structural mutation such as short insertions, deletions, and copy number variations.

Considering the cost and effectiveness of genotyping scales and cohort sizes, most of the population-based association studies choose genotyping screening methods, such as DNA microarrays. These methods measure thousands to millions of known SNPs in well-studied organisms, such as humans and mice. The targeted polymorphisms depend on the chip designs. For example, Immunochip contains 196,524 polymorphisms (718 InDels and 195,806 SNPs) on most reported loci involved in autoimmune and inflammatory diseases (20), whereas other custom genotyping chips contain loci designed for specific research areas, such as Metabochip (21) or cardiovascular disease chip (22). The number of variants that can be detected using genotyping chips has increased over the years, but even the high-density 5 million SNPs chip (Illumina OMNI5) covers only a small fraction of the 3.3 billion bases in the human genome.

In order to improve the power in discovering genetic associations on the regions poorly covered by DNA microarrays, genotype imputation approaches are often used to expand the coverage. For example, a commonly used genetic imputation server (https://imputationserver.sph.umich.edu/index.html#)! takes the ~60,000 public available human haplotypes, covering ~40,000,000 SNPs, as a reference to impute millions of missing SNPs based on the measured genotypes and linkage disequilibrium (LD) structures (23).

Before association analysis, genotype data should pass a standard quality control (QC) at both individual level and SNP levels. Individuals with discordant sex information, outlying missing genotype or heterozygosity rate should be excluded

TABLE 1 | Typical approaches in omics measurements.

	sequencing-based	microarray-based	others
genetics	whole-genome-seq, whole-exome-seq	Illumina OMNI5, Immunochip etc.	–
epigenetics	ATAC-seq, whole-genome bisulfite-seq, RRBS-seq, DNase-seq, FAIRE-seq, ChIP-seq, etc.	MethylationEPIC BeadChip, ChIP-chip, etc.	–
3D chromosome	Hi-C, etc.	–	–
gene expression	RNA-seq, scRNA-seq, SLAM-seq	Affymetrix Genome U133 array, Illumina Whole-Genome Gene Expression BeadChips, etc.	–
protein level	–	–	Immunoassay, MS -based approaches
metabolites	–	–	NMR, MS-based approaches
microbiome	16s rRNA-seq, metagenomics, etc.	–	–
cellomics	single cell sequencing approaches	–	FCM, CyTOF

(24). Duplicates and relatives could be identified by calculating identity by descent (IBD), and a multi-dimensional scaling plot merging with reference data such as the 1000 Genomes project (25) could help with the identification of individuals of divergent ancestry. SNPs failed in genotyping and/or imputation and SNPs with low frequency and/or that deviate from the Hardy-Weinberg equilibrium are commonly removed before association analysis, especially in array-based studies, because those signals usually relate to bad genotyping quality. However, some SNPs with low frequency may also contribute to rare diseases or phenotypes. With the increase in genotyping quality, more and more recent studies focus on the function of rare alleles (minor allele frequency [MAF] < 0.01) (26–29).

EPIGENOME AND 3D CHROMOSOME MEASUREMENT

Epigenetics describes the study of chromatin traits (either in DNA or histones) that do not involve changes in the nucleotide sequence. Epigenetics measurements are mainly characterized by the changes in histone modification (methylation and acetylation), DNA methylation, chromatin modification, chromatin accessibility, and chromosome structure.

DNA methylation is the process of adding methyl groups to DNA molecules, almost exclusively in CpG dinucleotides with the cytosines on both strands being methylated. This process usually acts in promoter regions to repress gene transcription, and abnormal hypermethylation, which results in transcriptional silencing, is often associated with immune diseases or used as a biomarker (30). Genome-wide techniques, such as whole-genome bisulfite sequencing (WGBS) (31), reduce representation bisulfite sequencing (RRBS-seq) (32) and other non-targeted DNA methylation profiles, provide an opportunity to discover novel biomarkers. Other techniques, such as bisulfite-amplicon sequencing (BSAS) (33) and methylation arrays (34), detect the methylation status of CpG dinucleotides. Similar to genotyping arrays, the targeted regions from methylation arrays are based on the chip design. For the study of the human immune system, some well-established arrays can provide comprehensive coverage. For example, MethylationEPIC BeadChip covers over 850,000 methylation sites, making it ideal for an epigenome association study within big cohorts (35).

As the essential proteins that pack and order the DNA into structural units, histones play a role in gene regulation (36). Histone modification describes the post-translational modifications of histones, including methylation, acetylation and others. Histone methylation often occurs as arginine (R), lysine (K), or histidine (H) residues of histone H3 or H4 being monomethylated (me1), demethylated (me2), or trimethylated (me3). Array-based and sequencing-based approaches, such as ChIP-chip and ChIP-seq (37), are used to identify specific histone modifications that bind to DNA regions or domains.

Chromatin modifications and accessibility is another important aspect of epigenetic changes. One of the most widely-used techniques to capture chromatin accessibility is called Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq). A standard “bulk” ATAC-seq measurement detects genome-wide open chromatin within a pooled sample or tissue, while in order to capture cellular heterogeneity, single-cell ATAC-seq measures chromatin accessibility in thousands of individual cells, which can generate genome-wide profiles from 10k to 100k cells per experiment (38). Alternative techniques are also used to investigate chromatin phenomena, such as DNase-seq and FAIRE-seq, which measure open chromatin in regulatory regions, MNase, which identifies well-positioned nucleosomes, and ChIP-seq, which is used to detect binding sites of specific transcription factors (39).

Most epigenetic measurements also come with technical errors and biases. Biological replicates and technical replicates can help to characterize variability between samples and sequencing runs. Putting replicates of different conditions in the same batch is also important to avoid batch effects confounding treatment effects. Large projects, such as the Encyclopedia of DNA Elements (ENCODE), have provided standard pipelines for processing many types of epigenetic data, such as ChIP-seq and ATAC-seq. However, this is not applicable in all cases. Applying appropriate QC strategies and software that accounts for bias effects according to the experiment design is essential to obtain robust results. To increase the coverage of epigenetic measurements, several methods, such as ChromImpute (40), Melissa (41), Avocado (42), and SCALE (43), provide imputation approaches for different epigenetic markers. However, the existing imputation approaches have several limitations (16), and are not as widely applied as genotype imputation methods.

3D chromosome structure describes how chromosomes are folded, packaged, and organized into functional compartments, and how different compartments are interconnected. Orthogonal ligation-based approaches include DNA-FISH, which can help with nuclear architecture visualization, and chromosome conformation capture (3C) techniques. One of the 3C techniques, Hi-C, is the most widely used approach to detect interactions between different genome regions (in gigabase-scales) (39, 44). Single-cell adaptation of Hi-C methods are also used to investigate the interactions in individual cells (45).

Ligation-based approaches have the limitation of detecting DNA fragments connected with multiple genomic regions. To overcome this limitation, orthogonal ligation-free methods including genome architecture mapping (GAM) (46), split pool recognition of interactions by tag extension (SPRITE) (47) and chromatin-interaction analysis *via* droplet-based and barcode-linked sequencing (ChIA-Drop) (48) were developed.

TRANSCRIPTOME MEASUREMENT

The transcriptome comprises all RNA molecules, both coding and non-coding transcripts, in a single or population of cells. Traditional qPCR techniques can only quantify a limited number of genes at the same time. The most commonly used high-throughput techniques are RNA sequencing (RNA-seq) and microarray, and they can detect a large number of genes. Similar to genotyping methods, a sequencing-based approach (RNA-seq) can quantify the entire transcriptome, while microarray-based approaches (e.g., Affymetrix Genome U133 array and Illumina Whole-Genome Gene Expression BeadChips) are designed to target most known genes. In addition, a typical RNA-seq can detect alternative splicing and rare isoforms, which microarray-based techniques cannot.

Certain coverage is required for sequencing data, which depends on the aim of the study. For instance, a bulk RNA-seq study for human differential expression profiling requires 10-25 million reads per sample, while alternative splicing or allele-specific expression analysis need 50-100 million, and identifying novel transcripts requires >100 million reads per sample.

However, a "bulk" like measurement of transcriptome cannot deal with the cell heterogeneity and can be influenced by cell composition changes. Single-cell RNA sequencing (scRNA-seq) was designed to uncover the transcriptome diversity in heterogeneous samples, characterizing the transcriptome in cell resolution. There are several approaches of scRNA-seq, among them are plate-based (Smart-seq2) (49) and droplet-based (10x Genomics) the most commonly used ones. Usually, as few as 10,000 to 50,000 reads per cell are enough to detect cell types, and 500,000 reads can cover most of the genes (50).

In order to increase exonic coverage and accuracy of gene quantification, polyA selection library preparation is commonly applied in scRNA-seq approaches such as 10x scRNA-seq (51). This will, however, miss the important immune repertoire profiling, such as B-cell and T-cell receptors, which is mainly distinguished by their 5' mRNA sequences. Thus, sequencing facilities, such as 10x genomics, provide full length paired B-cell and T-cell repertoire sequencing, simultaneously, when examining cellular gene expression level. Combined with transcription measurement, this information can improve our understanding of clonal expansion and better characterize immune cell heterogeneity and functions (52).

SLAM-seq detects the newly synthesized RNAs using a metabolic RNA labeling approach. Compared to the other scRNA-seq techniques, this method can track the transcriptome dynamics (53). For example, scSLAM-seq was applied to characterize the onset of infection with lytic cytomegalovirus in single mouse fibroblasts (54).

The transcriptome reflects the dynamic changes in biological processes, which is much more unstable. So, an appropriate sampling strategy on transcriptome data is crucial. In addition to the quality control, normalization is usually performed within a sample and between samples. When considering comparison analysis, it is also necessary to have biological replicates and check for batch effects using clustering-based approaches. There are many computational tools handling batch effects. Of note, integration approaches (55), as included in Seurat (56) and Harmony (57) packages, are commonly used in scRNA-seq analysis which detect the consistent cell type signals from different batches or measurements. However, when the batch difference is confounded with other group information, it will be tough to filter out the batch effects. In addition to batches, it is also important to consider other potential confounders in experiment design. For example, transcriptional differences were observed between males and females in COVID-19 patients (58), thus a gender-balanced design in a case-control study will lead to an unbiased conclusion for COVID-19. Moreover, when considering sampling tissues for immune responses, circulating leukocytes are often measured for systemic inflammatory responses, while inflamed tissues are measured for local inflammatory responses. In order to expand the capacity, deconvolution approaches have been applied to bulk RNA-seq data to characterize cell type compositions (59, 60), while expression recovery methods have been applied to single-cell RNA-seq data to reduce the dropout noise (61, 62). Like imputation approaches in genome and epigenome studies, one should be aware and careful with the potential false signals in these recovery or deconvolution approaches.

PROTEOME AND METABOLOME MEASUREMENT

Proteins are the major transcriptional products and functional units in the immune system. Immune molecules like immunoglobulins and cytokines are usually detected and/or quantified by immunoassays such as immunofluorescent staining, enzyme-linked immunosorbent assay (ELISA), enzyme multiplied immunoassay technique (EMIT), or mass spectrometry (MS)-based approaches.

In addition to independent measurement, proteins can be also measured together with RNA transcripts. CITE-seq provides an opportunity of identifying surface proteins along with RNA-seq. This approach is often used for cell labeling in scRNA-seq (63). Cells in different research groups (e.g., under different treatments, from different tissues) could be labeled with different antibodies as hashtags, then sequenced together as one pool. This process has two advantages: decreasing cost and excluding potential batch effects. In addition, as we also know that some immune cell types have specific cell markers, this approach can also be used to identify cell types. For example, the detection of CD3e, CD4 and CD8a proteins on the cell surface could help to distinguish CD4 T cells from CD8 T cells (64). Moreover, there is a new technique called INs-seq, which can measure intracellular protein activity along with scRNA-seq. This new technique shows a large potential of applications in immune-related studies (65).

The study of metabolic processes that regulate immune cell responses, which is referred to as immunometabolism, has become an exciting area in translational research, and is paving the way for novel therapies in immune-related diseases. The intermediate or end products of cellular metabolism are metabolites, which include, but are not limited to, lipids, fatty acids, amino acids, bile acids, and cholesterols. Considering the regulatory effects of metabolites on the immune response (12, 66, 67), the metabolome has become an important subject to study in immunological research.

Approaches to study the metabolome can be classified into targeted and non-targeted techniques. Nuclear magnetic resonance (NMR) spectroscopy is one of the most commonly used techniques, detecting specific nuclei in the target molecule (68). Compared to NMR, mass spectrometry (MS)-based approaches are more high-throughput and quantify metabolites in a non-targeted way, which detect mass-to-charge ratio (69). However, MS-based approaches have a limitation in annotating metabolites, which is the major drawback of this method in contrast to NMR. Metabolites data could be acquired from different sources of samples. Among them, circulating metabolites are the most commonly measured. There are also many studies about fecal and urine metabolites.

Similar to transcriptome analysis, a proper normalization (usually a log transformation) is required in both the proteome and metabolome data process. Secondly, biological replicates and batch effects have to be taken into consideration as well. In addition to linear regression, more advanced computational tools, such as ROIMCR (70), can also be used to reduce the batch effects and to identify metabolites that associate with immune responses. In terms of sampling tissues, in addition to blood cells and inflamed tissues, proteome and metabolome can also be measured in urine, which is thought to be a rich source but underestimated in recent studies (71–73). In addition, fecal metabolites are usually studied together with microbiota, which affects immune homeostasis and susceptibility of the host to immune-mediated diseases. Of note, there is a recent study reporting a reference map for serum metabolites (74), which can serve as a guide to control for irrelevant confounders in serum metabolite studies.

DIGESTIVE SYSTEM MICROBIOTA MEASUREMENT

Microbiota refer to all micro-organisms in a certain environment, for example the human digestive system. It has been reported to vary among individuals, to influence host immune functionality and to be involved in immune-mediated disease pathology (75–77). The commonly used approaches to study microbiota include 16s rRNA sequencing and metagenomics sequencing. After excluding host (human) reads, microbiota reads are aligned to the known microbiome genomes to identify the taxonomies and abundance. While there are also other omics approaches including metatranscriptomics, metaproteomics, and metabolomics, which target transcripts, proteins, or metabolites from microbiota (78).

Of note, studies on human microbiota usually have relatively low concordance compared to other omics data studies. A recent study has reported a number of host variables that could confound human gut microbiota researches. To be exact, body mass index (BMI), sex, age, geographical location, alcohol consumption, bowel movement quality (BMQ), and diet should be balanced in cases and controls when comparing gut microbiota compositions (79). In the context of sample collection, most of the microbiota samples are acquired from the stool, while urine and exhaled gas could be another important resource for microbiome detection (80, 81).

CELLOMICS MEASUREMENTS

Cellomics measurements often reveal the systemic responses at the level of cells and tissues, typically including cell composition, cellular localization and trafficking analyses. Cell composition is measured as cell type abundance or proportion, which is commonly quantified by flow and mass cytometry (82) (FCM and CyTOF) or single-cell sequencing techniques. With the help of cell surface markers or cellular-specific expression markers, both techniques can characterize hundreds of circulating cell subpopulations covering major immune cells involved in innate and adaptive immune responses (i.e., neutrophils, monocytes, lymphocytes, and their subtypes). Additionally, high-content screening (HCS) is commonly used to track cellular changes, including their localization, trafficking and morphologic phenotypes (83, 84).

SYSTEMS ANALYSIS ON OMICS DATA

After data collection and pre-processing with appropriate strategies, the next big challenge lies in linking different omics datasets and clinical phenotypes. For a certain trait or disease, a systems model can be built to specify the role and effect of different data layers. In this model, the qualitative or quantitative characteristics are linked by their relationships, which need to be estimated *via* comparison, association and other systems approaches. These links can simply be a correlation, or a regulatory or causal effect. In this section, we introduce general system approaches among different omics data and provide representative examples of how they can be applied in immunological studies.

GENOME-WIDE ASSOCIATION ANALYSIS AND QUANTITATIVE TRAIT LOCUS MAPPING

Genome-wide association studies (GWAS) aim to scan the whole genome to find genetic determinants of certain traits. When considering a binary trait (e.g., case-control), we compare allele frequency in two groups of individuals, for example one disease group and one healthy group. A chi-squared test is often applied to test for statistical significance. It is usually considered that there are ~1,000,000 independent loci in the human genome, so a p-value less than the Bonferroni corrected threshold of 0.05/1,000,000 (5×10^{-8}) is regarded to be genome-wide significant (85).

To date, GWAS have identified ~5000 genetic risk loci of immune-related diseases in ~400 studies (86). Those findings improved our understanding of genetic factors influencing immune-mediated diseases, further pointing to the genetic basis of pathology as well as treatment targets.

Generally, GWAS identify pathogenetic genetic factors contributing to phenotypes (diseases), though those variants will not cause disease directly but affect intermediate molecules. Quantitative trait locus (QTL) analysis is a statistical method to discover the genetic basis of the intermediated phenotypes, such as gene expression (eQTL) (87), splicing (sQTL) (88), metabolites (mQTL) (29), methylation (meQTL) (89, 90), and immune traits (91, 92).

After data normalization, a linear regression between each genetic variant and each quantitative trait is applied. Covariates are crucial aspects of the regression model of QTL analysis. Based on the type of omics, different covariates should be included in the model to correct the detected phenotypes. In general, basic host features such as age and sex are considered, and a population structure has to be additionally taken into account, especially in large cohorts with samples from admixed ancestry (93, 94).

eQTLs are the associations between SNPs and expression of genes, which provide insights of the function of genetic variants. eQTLs can explain 10% - 50% heritability of a phenotype/disease (95), which means that gene expression variation is one of the major consequences of genetic variants. It is very useful for prioritizing pathogenic genes when there is an association between a gene expression and a pathogenic genetic variant. Based on the position, eQTLs are classified into cis-eQTL (eQTL within 1Mb of the gene) and trans-eQTL (eQTL located outside 1Mb of the gene). Among them, trans-eQTLs are more tissue-specific than cis-eQTLs (88). Of note, tissue-specific eQTLs provide a way for prioritizing pathogenic tissues (96).

QTL analysis on epigenome identifies the associations between genetic variants and epigenetic modification. Most genome-wide significant disease-associated loci (~93%) are located in non-coding regions (97), particularly, regulatory elements identified by ENCODE (98) and Roadmap projects (99). These observations highlight the importance of epigenome in the genetic regulation of diseases and immune functionality. Similar to eQTL analysis, this analysis could help us find the potential epigenetic mechanism responsible for the association between genetic variants and immune traits/diseases. For example, a study investigated genetic variants that affect the activity of cis-regulatory domains (aCRD-QTLs) or correlation structure within cis-regulatory domains (sCRD-QTLs) in 317 lymphoblastoid and 78 fibroblast cell lines, and their consequence on gene expression (100). At the same time, genetic variants can also affect methylation (meQTL) by influencing the binding of DNA methyltransferase (DNA MTase). Large meQTL studies in blood samples showed significant enrichment in autoimmune diseases such as ulcerative colitis and Crohn's disease (101).

pQTL mirrors the associations between genetic variants and protein level. About 40% of cis-protein quantitative trait loci (pQTLs) are also eQTLs, as expected, indicating a sequential genetic regulation between gene expression level and protein levels. By applying pQTL analysis, we could identify the potential mechanism, at the protein expression level, behind the association from genetic variants to immune-related phenotypes. Same as with cis-eQTLs, cis-pQTLs are also located around transcription start sites (TSS). Notably, pQTL showed a significant enrichment on missense, 3UTR and splice region (102). pQTLs could also help with prioritizing causal proteins/genes of immune traits/diseases. For example, a pQTL of serum IL18R1 and IL1RL1 also associates with atopic dermatitis. This association between genetic locus and protein level indicates a possible involvement of IL18R1 and IL1RL1 in atopic dermatitis pathology (102).

Metabolites that mediate the association between genetic variants to immune functionality and immune diseases could be discovered in an mQTL analysis. More than 140 genomic loci are associated with circulating metabolite features explaining a median 6.9% heritability (103). Overlaps between mQTLs and immune traits QTLs suggest the role of metabolic processes in the genetic regulation of immune functionality. For instance, a mQTL study indicates that mQTL loci ARHGEF3 (rs1354034) and LRRC8A (rs13297295) also affect platelet function and neutrophil function, respectively (104).

Immune phenotypes such as circulating immune cell proportion and cytokine production capacity in response to stimulations are crucial parameters when characterizing immune activities. Understanding the genetic determinants of immune phenotypes can provide insights into immune function and immune-mediated diseases. A human functional genomics project has identified >20 genetic factors determining immune cell proportions and cytokine production upon stimulations, which provided a link between genetic control and inter-individual variation (92, 105).

INTEGRATION OF MULTIPLE GENETIC ASSOCIATION PROFILES

In the context of immunological research, multiple diseases, and molecular and cellular phenotypes can be regulated by the same genetic factors, indicating an internal association between them. Integration with multiple genetic profiles can provide insights and build connections between associated phenotypes. Ideally, such genetic profiles can be directly built from GWAS and QTL analysis of different layers from the same individuals. Otherwise, they can be also collected from different population-based cohorts. A number of computational approaches have been developed to discover the link. In particular, approaches like colocalization (106), genetic correlation (107) and Mendelian randomization (MR) (108) take genetic variants as the instrumental variables to infer the association or causality when multiple traits are associated with the same locus.

Colocalization analysis evaluates the association from each of the single locus, and it helps to identify the phenotypes that share the same genetic regulation. Examples of colocalization analysis include a study integrating genetics, epigenetics and transcription to identify colocalization of molecular traits from CD14+ monocytes, CD16+ neutrophils and naïve CD4+ T cells

(109). Results from this analysis illustrate molecular mechanism at autoimmune diseases-associated variants, including an alternative splicing signal around SP140 in T cells which might be involved in Crohn's disease pathology.

Genetic correlation considers the full summary statistics to describe to which extent the genetic background is shared between two phenotypes. An example from a LD regression-based genetic correlation approach showed a shared genetic basis of autoimmune diseases such as Celiac disease and type 1 diabetes (107). This indicates a similar pathological mechanism between these two diseases.

MR is a statistical method working on the step from association to causality. If one trait (exposure) is causal to another trait (outcome), then the genetic factors contributing to the exposure should also contribute to the outcome. This would be reflected in the correlation between effect sizes of the same genetic variant on exposure and outcome. There are many examples of immune-related studies that applied MR, which led to the identification of causal relationships between IL-6 signaling and rheumatoid arthritis (110), IL-18 and inflammatory bowel disease (111) and between eosinophilic indices and asthma (26).

COMPARISON AND ASSOCIATION OF EPIGENOME AND 3D CHROMOSOME STRUCTURES

Systems analysis of epigenetic changes can investigate their influence on and changes induced by immune functionality or variation as well as disease susceptibility and development (112, 113). As an example, the impact of cytokines was studied on the epigenome of insulin releasing cells (β cells) from type 1 diabetes pancreases. By measuring ATAC-seq, Chip-seq and RNA-seq, the authors identified proinflammatory cytokines induced neo/primed epigenetic events in human β cells (114). Moreover, in immune systems, the effects of epigenetic changes lead to long-term alterations in the metabolic and transcriptional pathways, and further induce immune memory (115) or immunological diseases (116). Thus, epigenomics is another vital area for better understanding of the personalized immune system.

While genetics is stable, the epigenome is subject to dynamic changes, which can be induced or affected by host and environmental factors, such as smoking, drug usage, diet, aging, inflammation, disease, and exposure to pets. Considering that epigenetic changes affect gene transcription levels, the epigenome is a pivotal part to study when trying to understand immunological networks.

In a case-control study, differential accessible regions (DARs) could be identified in an ATAC-seq data, as well as differential methylation positions/regions (DMP/DMRs) in bisulfite sequencing and methylation array. Instead of comparison analysis, association analysis is applied to continuous phenotypes to get associated regions. Upon the position of acquired regions, we could further map them to the corresponding genes. More specifically, by checking which gene TSS regions are overlapped with the peaks/regions, the peaks/regions could be matched to genes, and then for pathway analysis to get more biological meanings. For example, in a multi-omics study on mixed-phenotype acute leukemia, researchers associated scATAC-seq with transcription responses from scRNA-seq and antibody captured from CITE-seq. Despite widespread epigenetic heterogeneity of chromatin accessibility within patients, they reported common malignant signatures across patients, and thus revealed both distinct and shared molecular mechanisms of mixed-phenotype acute leukemia (117).

Another application of epigenetic analysis is to annotate the function of the identified regions, based on the signals from epigenetic markers. A tool (118) used a multivariate hidden Markov model applied to annotate regulatory elements (e.g., Transcription starting sites, enhancers, promoters) with histone markers (e.g. H3K4me1, H3K4me3, H3K27me3, H3K9me3, H3K36me3) binds to the chromosomes. Applying this method, an example learnt the chromatin states in mice and humans, and reported the up-regulation of immune regulatory regions in Alzheimer's disease (119).

The analyses on 3D chromosomes are generally similar. In a case-control study with Hi-C data, we could get the compartment switches in a comparative analysis. We could further predict the interactions between those segments (120). Referring public epigenetic databases or genome annotations, we could check the overlap between switched compartments or interactions and known epigenetic markers or elements. Based on this information, we could again associate the changes with other immune profiles or annotate the involved regulatory elements. For example, in a study of lineage commitment of early T cells with Hi-C data, authors found wide compartment re-organizations across chromosomes from a transition between T cell double-negative-2 stage to double-negative-3 stage, and later double-negative-4 stage to double-positive stage. They annotated the changes with domain scores, and more interestingly, they found the changes in the domain scores between the two transitions are positively correlated, which suggests the re-organization at the former transition is actually reinforced at the later transition (121). Another example includes a study on activated T cells, that identified activation-sensitive interactions related to autoimmune diseases captured by Hi-C data (122).

To capture the changes that occur in cellular activation and differentiation, time-series study is another hot topic in associating epigenome and 3D chromosome structures to immune responses. For example, a recent study elucidates the chromosome conformational changes in B lymphocytes as they differentiate and expand from a naive, quiescent state into antibody secreting plasma cells (123). The authors reveal that the changes to 3D chromatin structure occur in two discrete windows, associated with prolonged time in the G1 phase of the cell cycle. Their results also suggest chromosome reconfiguration is linked to a gene expression program that controls the differentiation process required for the generation of immunity.

COMPARISON AND ASSOCIATION OF TRANSCRIPTOME AND PROTEOME

As the downstream products of genetic and epigenetic regulation, transcriptome and proteome changes directly reflect the influence of genetic and epigenetic variants. Comparison and association studies of transcriptome and proteome have allowed researchers to estimate functional units and validate hypotheses in immune regulation.

As for a case-control study, the first and direct analysis is identifying differentially expressed genes/proteins (DEGs/DEPs), followed by pathway analysis. If the corresponding phenotypes are continuous, then associated genes/proteins will be identified before pathway analysis. Examples include many transcriptome/proteome studies upon the severe infectious disease COVID-19. Transcriptome measurement across samples from healthy, moderate patients and severe patients suggests an overall acute inflammatory response in COVID-19 patients, whereas transcriptional responses of high cytotoxic effector T cells are associated with moderate patients, and deranged interferon responses are associated with severe patients (124). Moreover, a urine proteome study identified 1986 urine proteins showing significant level changes in COVID-19 patients than in healthy controls (125).

Different from bulk RNA-seq, the adding information in scRNA-seq: cell composition, provides more analysis potentials. In a case-control study, in addition to DEGs and enriched pathways identification within each cell cluster/type, cell proportion could be compared between groups while novel cell subpopulation could also be identified in particular cases. For example, a scRNA-seq on two COVID-19 cohorts reported identical dysfunctional neutrophil clusters in severe patients' blood (9). When considering the TCR/BCR analysis, it would be interesting to explore the clonotype expansion and diversity under different conditions (126, 127), immune development stages (52), or antigen specificity (128). Usually, a clonal expansion means an adaptive immune response targeting certain stimulation, since a certain receptor is the mediator of specific antigen recognition.

Since transcriptome/proteome data is rapidly responding to environmental changes, with the transcriptome/proteome analysis in a time-series study, we could associate the dynamics with infection or stimulation to comprehensively understand the host immune responses. A nice example is demonstrated in a study of influenza vaccination efficiency, where authors measured the hemagglutination-inhibition (HAI) antibody titers and transcriptional responses at baseline and multiple time points post-vaccination. By comparing the profiles between day 28 and day 180, the authors describe individual categories as temporary and persistent responders and illustrate the underneath molecular mechanism (129). Many approaches have been developed for time-series studies, such as regression-based method like maSigPro (130) and a fusion method like O2-PLS (131). Of note, the dynamic study can also be achieved by applying a trajectory analysis such as pseudotime analysis (132, 133) and RNA velocity analysis (134) in scRNA-seq analyses. In a recent study on COVID-19, researchers longitudinally measured samples at several time-point after symptoms, and applied pseudo-time trajectory inference on scRNA-seq data of epithelial cells from the upper respiratory tract. Based on the trajectory, they predicted a new, alternative differentiation pathway that is dependent on the interferon response and marked by interferon-stimulated genes, such as *ISG15*, *IFIT1*, and *CXCL10* (135).

Co-expression analysis among transcriptome or proteome provided information about gene co-regulation and interactions. These co-expression relationships are inferred by different association methods, such as a weighted gene co-expression network analysis (WGCNA) (136) applied on transcriptome to identify consistent expression patterns among genes. The identified associations among gene expression could be applied to predict gene co-regulatory networks, further to prioritize genes involved in the same pathways (137). At protein level, parts of these co-expression relationships could further be explained by protein-protein interactions, which are also collected by several protein-protein interaction databases, including the innateDB (138) who particularly focus on immune interactions. In application, similar to gene co-expression networks, protein-protein interaction relationships could help with functional/pathway enrichment analysis (139).

In the recent single-cell experiments, the co-expression relationships are further applied to predict the cell-cell interactions. By detecting the correlation between known ligand and receptor genes among different cell sub clusters, we could infer the potential communications between cell populations (140). This analysis fits well with immune network analysis. For example, by detecting ligand and receptor genes signals, a recent study identified cross-talks between CD8+ T cells and epithelial cells altered in the colon of ulcerative colitis patients compared to healthy controls (141). Additional methods, such as NicheNet (142), also take knowledge of gene regulatory networks or protein-protein interaction networks from public databases and literatures, then build a model to further predict the activated targets of the cell-cell interactions by correlating the ligands expression level with its potential downstream gene or protein level interactions. In an example study of cell-cell interaction underlying the tissue-specific imprinting of macrophages, the authors deciphered the interaction signals driving monocyte recruitment, engraftment, and acquisition of the Kupffer cells associated transcription factors, and they identified the contributions of different cells to Kupffer cell niche (143).

COMPARISON AND ASSOCIATION ON METABOLOME/MICROBIOME

Metabolome or microbiome are additional factors that reflect, or affect, a person's state of health (144, 145). Similar to transcriptome or proteome, comparison and association analysis could be applied on metabolome and microbiome data. However, metabolome can be hardly linked to genes, which leads to different strategies of interpretation. Taking KEGG (146) and HMDB (147) as references, an online tool MetaboAnalyst performed metabolic pathway enrichment and network analysis on the identified metabolites (148). An example serum study on COVID-19 detected accumulation of 11 steroid hormones and suppression of amino acid metabolism in patients (8)

As for the gut microbiome, a diversity analysis could be applied to taxonomy data. There are different strategies available for functional profiling on the gut microbiome data. For example, HUMAnN takes metagenomic or metatranscriptomic sequencing data as input to identify gene families and abundances (149). Gene families could be further matched to broader functional categories, such as MetaCyc metabolic pathways and GO categories for functional interpretation. For example, a study associated gut microbiome features to cytokine production capacity, and found microbial metabolic pathways: palmitoleic acid metabolism and tryptophan degradation to tryptophol showed associations with TNFα and IFNγ production (150).

As in transcriptome and proteome analyses, time-series studies could provide valuable information in metabolome and microbiome data. For example, in a study of metabolic functions of gut microbes from patients with Inflammatory Bowel Diseases, fecal samples were collected at baseline and 2, 6, and 14/30 weeks after induction of therapy to collect metabolic and microbiota profiles. The observed association in dynamics of metabolites and diversity shifts of microbiota reveals the heterogeneity of the disease, and helps the authors to build a silico model that might be used to identify patients likely to achieve clinical remission from the therapy (151).

INTEGRATION OF EPIGENOME, TRANSCRIPTOME, PROTEOME, METABOLOME, MICROBIOTA AND CELLOMICS

Besides associations between omics data and genetics, a simple association analysis between two different non-genetic omics data could be applied to the data measured in the same cohort with a large sample size to find the co-regulations behind (**Table 2**). For instance, eQTMs (associations between methylation and gene expression) provide a resource to integrate methylation and gene expression. Highly methylation can block the binding of transcription factors on promoters and enhancers. In line with expectation, most eQTMs showed negative correlations between methylation and gene expression, and negatively correlated eQTMs are enriched in active TSS regions (152). For another example, a study carefully characterizes the changes in the gut microbiota of patients suffering inflammatory bowel diseases and the interplay between microbiome composition and gut metabolites (153).

In the situation of a more complex multi-omics integration, more advanced technique like building multivariable regression model could take features from different omics to evaluate the accumulative effects/prediction power on a certain phenotype. An example study integrates genomic, metagenomic, metabolomic, immune cell composition, hormone levels and platelet activation profiles with cytokine response profiles in a population-based cohort. Results from multivariable linear regression and machine learning approaches such as elastic net show the accumulative contribution and predict power of genetic and non-genetic factors on cytokine response (154).

On the other hand, if the sample size is not allowed for association analysis, it might be applicable to check the intersections between the findings from different omics. For example, we could easily compare the regions identified in ATAC-seq, methylation array and Hi-C data. In addition, by matching a DAR to genes, and intersecting with DEGs, we could further check whether an epigenetic change has the potential in regulating gene expression.

DISCUSSION AND PERSPECTIVES

In this review, we have discussed the multi-omics application for immunological studies, from measurements and analysis to comparison or association of several typical layers (**Figure 2**). For system studies – in particular newly discovered infectious diseases or rare diseases with fewer prior knowledge – the choice

TABLE 2 | System analysis between omics.

	binary traits	epigenetics	gene expression	protein level	metabolites	microbiome	cellomics
genetics	GWAS	meQTL, CRDQTLs	eQTL, sQTL	pQTL	mQTL	mbQTL	cell proportion QTLs
epigenetics	DMRs/DARs/Compartment Switches/ Gained or lost Interactions	position-based overlap	gene-based overlap/ association	gene-based overlap/ association	association	association	association
gene expresion	DEGs	–	co-expression	gene-based overlap/ association	association	association	association
protein level	DEPs	–	–	coexpression/ interaction	–	association	association
metabolites	different abundance	–	–	–	association	association	association
microbiome	different composition	–	–	–	–	association	–
cellomics	Different cell composition, etc.	–	–	–	–	–	association

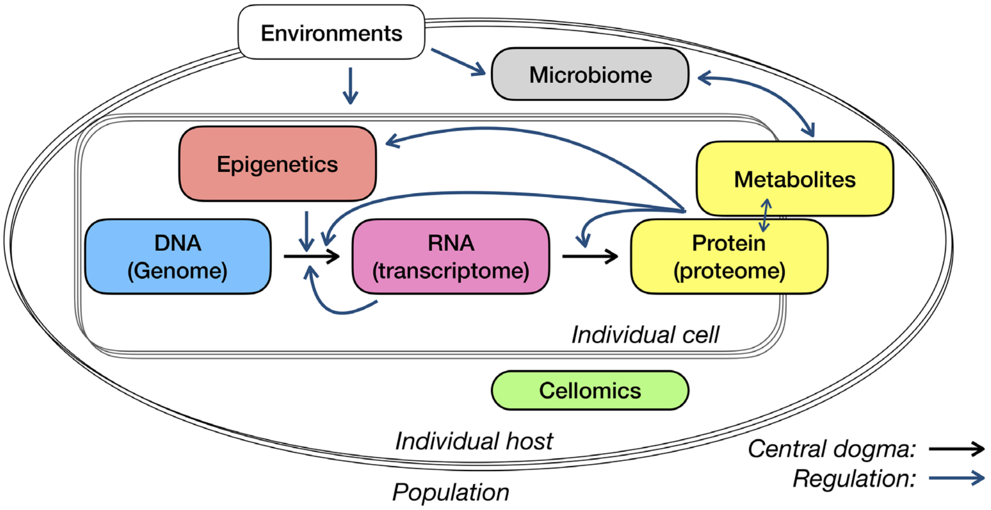


FIGURE 2 | Central dogma and regulations of different omics layers.

of data layers to collect and the selection of measuring approaches, target or non-target technique, bulk or single-cell level, can be as important as the analysis models and algorithms. Here, we discuss a few points that need specific attention in study design and interpretation, and subjects may see progress in the next few years.

There are some commonly used strategies of interpreting genetic associations. As the starting point of the central dogma of molecular biology (**Figure 2**), genetics has so far received a lot of attention and was associated with many types of data or phenotypes. In the interpretation of genome-wide associated loci, genes around them have also been regarded as the necessary and most essential compartments. The strategy to properly link loci with affected genes so far has been addressed on the position and associations between gene expression and genetic variants (i.e., eQTLs). In addition, functional annotation on identified loci, such as whether the variants are located on the regulatory elements or affected protein structure, may provide additional clues for loci interpretation in particular cases. Nevertheless, there are existing debates upon several aspects, for example, whether the host genome could influence the gut microbiome. It will never be nitpicking to be very careful with interpreting your microbiome QTLs.

Epigenetic could be used as a window to study environmental influence. In contrast to genetics, epigenetics often links the external factors to immune phenotypes. This is particularly true when considering the external effects as a risk to immune diseases, for example, smoking to asthma, because epigenetic modifications, such as methylation, are usually related to environmental exposures. Considering the various kinds of epigenetic changes, multiple types of epigenetic data are commonly used in one study and they often validate and complement each other. For example, an active TSS region could be identified by low methylation as well as high DNA accessibility (155), and the enhancer involved in a neo chromatin interaction identified in Hi-C data could be characterized as a neo opening region in ATAC data (156). Considering the functional relationships, epigenetic data is commonly integrated with gene expression measurements. As the direct consequence of epigenetic modification, alteration in corresponding gene expression could be the best validation of the importance of your epigenetic studies.

scRNA-seq is usually applied together with Cellomics measurements. A cell composition discovered in scRNA-seq data could be validated with FCM-based approaches. FACS is also commonly used as a pre-filtering step to help with concentrating target cell types for scRNA-seq analysis. Especially, for the rare cell types (e.g., T regs in PBMCs), a pre-sorting process is necessary for concentrating on cells of interest.

Proteome, metabolome showing downstream immune functions require more attentions. As the downstream products of gene expression, protein or metabolites level measurements are not as popular as transcriptome measurements in current studies. This might because gene expression analysis takes advantage of the efficiency of next-generation sequencing and well-established microarray chips. Thus, there appears to be much room for further studies on proteome and metabolome in immune studies.

Proper measurement techniques and sampling tissues are crucial in an omics study. When considering the purpose of measurements, it is often appropriate to apply high-throughput and/or non-target approaches at the discovery stage, while single and/or target approaches are more commonly used for validation. Besides, except genome, all the other omics have tissue specificity. Data from the same tissue are more commonly associated. For example, associations between omics from blood samples could be easily interpreted, but it would be tricky and needs more biological basis to associate blood features with gut features.

A straight-forward joint visualization of multi-omics data is another challenge to better present and understand the interconnections across molecular layers as well as to fully utilizing the increasingly available omics data. Integrated tools or platforms that combined a comprehensive analysis workflow and interactive visualizations were often more preferable to researchers. Some examples are: PaintOmics3 (157) and Metascape (158), which provide powerful online frameworks for the multi-omics pathway analysis and visualization; Seurat (56), which focuses on analysis and visualization of single-cell omics data and supports easy connections to other popular analysis tools; and Omics Playground (159), who provides a user-friendly and interactive self-service bioinformatics platform for analysis, visualization and interpretation of transcriptomics and proteomics data. Moreover, trials of combing data sharing and interactive visualization along with research publication have also been made to improve the data dissemination. For example, by accessing to Immgen (160), FastGenomics (161) or DeCovid (58), researchers can explore and visualize their interested immune signatures on the COVID-19 datasets, which significantly increases impact of the studies.

To fully elucidate the biological processes involved in the immune system, several aspects remain unknown in omics studies. Firstly, due to sample accessibility, fewer studies have been performed on tissues other than blood. Taking meQTLs as an example, several big studies have been carried out blood samples (101, 162, 163). However, there are very limited sample size and/or studies about meQTLs in other tissues (164). Secondly, considering the high dynamics, rapid response and spatial specificity of the immune system, temporal and spatial studies can provide more insights into the dynamic process and spatial heterogeneity in immune activities and/or immune-related diseases etiology. For example, the process that immune cells are activated by interacting physically and chemically with synapses is highly dynamic and depends on the spatial position of immune cells, neurons and glial cells. Despite its importance in immune functionality and immune-mediated diseases, our current knowledge is not sufficiently advanced, which calls for more comprehensive studies (165–167). Thirdly, as for population-based studies, there are much more of them in healthy individuals of European ancestry, while the studies in under-represented populations as well as in patients appeal for greater attention.

Considering the complexity of our immune system and patient heterogeneity, in terms of severity or treatment responses, for many immune-related diseases, the generation of personalized medicine is one of the most significant goals we can achieve through multi-omics studies (168). Personalized medicine stratifies a heterogeneous group of patients based on certain characteristics and provides treatment based on this stratification. In the case of infectious diseases, one of the personalized medicine trials is now being conducted for the treatment of sepsis using immunomodulatory interventions after stratification based on biomarkers identifying immunosuppression or hyper inflammation (169). In the field of tuberculosis, advances are being made too, as a clinical trial is now ongoing where tuberculous meningitis patients are being stratified based on genotype prior to treatment (170).

In conclusion, we systematically review measurements and analyses can be applied in immunological studies, which provide insights for personalized medicine. Through the development of high throughput techniques, e.g. single-cell RNA sequencing and mass cytometry, we now possess the tools to unravel the many complexities of the immune system in health and immune-related diseases, including infectious diseases, allergies and auto-immune diseases. With unbiased measurements and effective integration, multi-omics studies can help us understand the immune system and could lead to the development of personalized medicine.

AUTHOR CONTRIBUTIONS

XC made the conception and design of the review. XC and BZ drafted the manuscripts, supervised by YL. All authors contributed to the article and approved the submitted version.

FUNDING

XC was supported by the China Scholarship Council (201706040081). YL was supported by an ERC Starting Grant (948207) and the Radboud University Medical Centre Hypatia Grant (2018) for Scientific Research.

REFERENCES

1. Topol EJ, Califf RM. *Textbook of cardiovascular medicine*. Philadelphia, USA: Lippincott Williams & Wilkins (2007).
2. Atkinson MA, Eisenbarth GS, Michels AW. Type 1 Diabetes. *Lancet* (2014) 383(9911):69–82. doi: 10.1016/S0140-6736(13)60591-7
3. Reyes M, Filbin MR, Bhattacharyya RP, Billman K, Eisenhaure T, Hung DT, et al. An Immune-Cell Signature of Bacterial Sepsis. *Nat Med* (2020) 26 (3):333–40. doi: 10.1038/s41591-020-0752-4
4. Osterholm MT, Kelley NS, Sommer A, Belongia EA. Efficacy and Effectiveness of Influenza Vaccines: A Systematic Review and Meta-Analysis. *Lancet Infect Dis* (2012) 12(1):36–44. doi: 10.1016/S1473-3099 (11)70295-X
5. Warren RB, Smith CH, Yiu ZZ, Ashcroft DM, Barker JN, Burden AD, et al. Differential Drug Survival of Biologic Therapies for the Treatment of Psoriasis: A Prospective Observational Cohort Study From the British Association of Dermatologists Biologic Interventions Register (Badbir). *J Invest Dermatol* (2015) 135(11):2632–40. doi: 10.1038/jid.2015.208
6. Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, et al. Genetic Mechanisms of Critical Illness in Covid-19. *Nature* (2020) 591.7848:92–98.. doi: 10.1101/2020.09.24.20200048
7. Xiong Y, Liu Y, Cao L, Wang D, Guo M, Jiang A, et al. Transcriptomic Characteristics of Bronchoalveolar Lavage Fluid and Peripheral Blood Mononuclear Cells in COVID-19 Patients. *Emerg Microbes infections* (2020) 9(1):761–70. doi: 10.1080/22221751.2020.1747363
8. Shen B, Yi X, Sun Y, Bi X, Du J, Zhang C, et al. Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell* (2020) 182(1):59–72.e15. doi: 10.1016/j.cell.2020.05.032
9. Schulte-Schrepping J, Reusch N, Paclik D, Baßler K, Schlickeiser S, Zhang B, et al. Severe COVID-19 is Marked by a Dysregulated Myeloid Cell Compartment. *Cell* (2020) 182(6):1419–40. doi: 10.1016/j.cell. 2020.08.001

10. Bernardes JP, Mishra N, Tran F, Bahmer T, Best L, Blase JI, et al. Longitudinal Multi-Omics Analyses Identify Responses of Megakaryocytes, Erythroid Cells, and Plasmablasts as Hallmarks of Severe Covid-19. *Immunity* (2020) 53(6):1296–314. doi: 10.1016/j.immuni.2020.11.017
11. Zuo T, Zhan H, Zhang F, Liu Q, Tso EY, Lui GC, et al. Alterations in Fecal Fungal Microbiome of Patients With COVID-19 During Time of Hospitalization Until Discharge. *Gastroenterology* (2020) 159(4):1302–10. doi: 10.1053/j.gastro.2020.06.048
12. Erkelens MN, Mebius RE. Retinoic Acid and Immune Homeostasis: A Balancing Act. *Trends Immunol* (2017) 38(3):168–80. doi: 10.1016/j.it.2016.12.006
13. Kriete A, Eils R. *Computational Systems Biology: From Molecular Mechanisms to Disease*. San Diego, USA: Academic Press (2013).
14. Mirza B, Wang W, Wang J, Choi H, Chung NC, Ping P. Machine Learning and Integrative Analysis of Biomedical Big Data. *Genes* (2019) 10(2):87. doi: 10.3390/genes10020087
15. Jaumot J, Bedia C, Tauler R. *Data Analysis for Omic Sciences: Methods and Applications*. Amsterdam: Elsevier (2018).
16. Song M, Greenbaum J, Luttrell IVJ, Zhou W, Wu C, Shen H, et al. A Review of Integrative Imputation for Multi-Omics Datasets. *Front Genet* (2020) 11:570255. doi: 10.3389/fgene.2020.570255
17. Eckhardt M, Hultquist JF, Kaake RM, Hüttenhain R, Krogan NJ. A Systems Approach to Infectious Disease. *Nat Rev Genet* (2020) 21:339–54. doi: 10.1038/s41576-020-0212-5.
18. Savola P, Martelius T, Kankainen M, Koski Y, Eldfors S, Huuhtanen J, et al. Somatic Mutations in T Cells as Possible Regulators of Immunodeficiency. *Blood* (2018) 132(Supplement 1):515–. doi: 10.1182/blood-2018-99-110757
19. Netea MG, Joosten LA, Li Y, Kumar V, Oosting M, Smeekens S, et al. Understanding Human Immune Function Using the Resources From the Human Functional Genomics Project. *Nat Med* (2016) 22(8):831–3. doi: 10.1038/nm.4140
20. Cortes A, Brown MA. Promise and Pitfalls of the Immunochip. *Arthritis Res Ther* (2011) 13(1):101. doi: 10.1186/ar3204
21. Voight BF, Kang HM, Ding J, Palmer CD, Sidore C, Chines PS, et al. The Metabochip, a Custom Genotyping Array for Genetic Studies of Metabolic, Cardiovascular, and Anthropometric Traits. *PloS Genet* (2012) 8(8): e1002793. doi: 10.1371/journal.pgen.1002793
22. Keating BJ, Tischfield S, Murray SS, Bhangale T, Price TS, Glessner JT, et al. Concept, Design and Implementation of a Cardiovascular Gene-Centric 50 K Snp Array for Large-Scale Genomic Association Studies. *PloS One* (2008) 3 (10):e3583. doi: 10.1371/journal.pone.0003583
23. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-Generation Genotype Imputation Service and Methods. *Nat Genet* (2016) 48 (10):1284–7. doi: 10.1038/ng.3656
24. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data Quality Control in Genetic Case-Control Association Studies. *Nat Protoc* (2010) 5(9):1564–73. doi: 10.1038/nprot.2010.116
25. Siva N. 1000 Genomes Project. *Nat Biotechnol* (2008) 26(3):256–. doi: 10.1038/nbt0308-256b
26. Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, et al. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* (2016) 167(5):1415–29.e19. doi: 10.1016/j.cell.2016.10.042
27. Emdin CA, Khera AV, Chaffin M, Klarin D, Natarajan P, Aragam K, et al. Analysis of Predicted Loss-of-Function Variants in UK Biobank Identifies Variants Protective for Disease. *Nat Commun* (2018) 9(1):1613. doi: 10.1038/s41467-018-03911-8.
28. Ferraro NM, Strober BJ, Einson J, Abell NS, Aguet F, Barbeira AN, et al. Transcriptomic Signatures Across Human Tissues Identify Functional Rare Genetic Variation. *Science* (2020) 369(6509):eaaz5900. doi: 10.1126/science.aaz5900
29. Long T, Hicks M, Yu H-C, Biggs WH, Kirkness EF, Menni C, et al. Whole-Genome Sequencing Identifies Common-to-Rare Variants Associated With Human Blood Metabolites. *Nat Genet* (2017) 49(4):568–78. doi: 10.1038/ng.3809
30. Lund G, Andersson L, Lauria M, Lindholm M, Fraga MF, Villar-Garea A, et al. Dna Methylation Polymorphisms Precede Any Histological Sign of Atherosclerosis in Mice Lacking Apolipoprotein E. *J Biol Chem* (2004) 279 (28):29147–54. doi: 10.1074/jbc.M403618200
31. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA Methylomes at Base Resolution Show Widespread Epigenomic Differences. *nature* (2009) 462(7271):315–22. doi: 10.1038/nature08514
32. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced Representation Bisulfite Sequencing for Comparative High-Resolution DNA Methylation Analysis. *Nucleic Acids Res* (2005) 33 (18):5868–77. doi: 10.1093/nar/gki901
33. Masser DR, Berg AS, Freeman WM. Focused, High Accuracy 5-Methylcytosine Quantitation With Base Resolution by Benchtop Next-Generation Sequencing. *Epigenet chromatin* (2013) 6(1):1–12. doi: 10.1186/1756-8935-6-33
34. Mallik S, Odom GJ, Gao Z, Gomez L, Chen X, Wang L. An Evaluation of Supervised Methods for Identifying Differentially Methylated Regions in Illumina Methylation Arrays. *Briefings Bioinf* (2019) 20(6):2224–35. doi: 10.1093/bib/bby085
35. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical Evaluation of the Illumina Methylationepic Beadchip Microarray for Whole-Genome Dna Methylation Profiling. *Genome Biol* (2016) 17(1):1–17. doi: 10.1186/s13059-016-1066-1
36. Greer EL, Shi Y. Histone Methylation: A Dynamic Mark in Health, Disease and Inheritance. *Nat Rev Genet* (2012) 13(5):343–57. doi: 10.1038/nrg3173
37. Ji H, Jiang H, Ma W, Johnson DS, Myers RM, Wong WH. An Integrated Software System for Analyzing Chip-Chip and Chip-Seq Data. *Nat Biotechnol* (2008) 26(11):1293–300. doi: 10.1038/nbt.1505
38. Chen H, Lareau C, Andreani T, Vinyard ME, Garcia SP, Clement K, et al. Assessment of Computational Methods for the Analysis of Single-Cell Atac-Seq Data. *Genome Biol* (2019) 20(1):1–25. doi: 10.1186/s13059-019-1854-5
39. Meyer CA, Liu XS. Identifying and Mitigating Bias in Next-Generation Sequencing Methods for Chromatin Biology. *Nat Rev Genet* (2014) 15 (11):709–21. doi: 10.1038/nrg3788
40. Ernst J, Kellis M. Large-Scale Imputation of Epigenomic Datasets for Systematic Annotation of Diverse Human Tissues. *Nat Biotechnol* (2015) 33(4):364–76. doi: 10.1038/nbt.3157
41. Kapourani C-A, Sanguinetti G. Melissa: Bayesian Clustering and Imputation of Single-Cell Methylomes. *Genome Biol* (2019) 20(1):1–15. doi: 10.1186/s13059-019-1665-8
42. Schreiber J, Durham T, Bilmes J, Noble WS. Avocado: A Multi-Scale Deep Tensor Factorization Method Learns a Latent Representation of the Human Epigenome. *Genome Biol* (2020) 21(1):1–18. doi: 10.1186/s13059-020-01977-6
43. Xiong L, Xu K, Tian K, Shao Y, Tang L, Gao G, et al. Scale Method for Single-Cell Atac-Seq Analysis *Via* Latent Feature Extraction. *Nat Commun* (2019) 10(1):1–10. doi: 10.1038/s41467-019-12630-7
44. Kempfer R, Pombo A. Methods for Mapping 3d Chromosome Architecture. *Nat Rev Genet* (2020) 21(4):207–26. doi: 10.1038/s41576-019-0195-2
45. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, et al. Single-Cell Hi-C Reveals Cell-to-Cell Variability in Chromosome Structure. *Nature* (2013) 502(7469):59–64. doi: 10.1038/nature12593
46. Beagrie RA, Scialdone A, Schueler M, Kraemer DC, Chotalia M, Xie SQ, et al. Complex Multi-Enhancer Contacts Captured by Genome Architecture Mapping. *Nature* (2017) 543(7646):519–24. doi: 10.1038/nature21411
47. Vangala P, Murphy R, Quinodoz SA, Gellatly K, McDonel P, Guttman M, et al. High-Resolution Mapping of Multiway Enhancer-Promoter Interactions Regulating Pathogen Detection. *Mol Cell* (2020) 80(2):359–73. doi: 10.1016/j.molcel.2020.09.005
48. Koch L. Getting the Drop on Chromatin Interaction. *Nat Rev Genet* (2019) 20(4):192–3. doi: 10.1038/s41576-019-0103-9
49. Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-Seq2 for Sensitive Full-Length Transcriptome Profiling in Single Cells. *Nat Methods* (2013) 10(11):1096–8. doi: 10.1038/nmeth.2639
50. Haque A, Engel J, Teichmann SA, Lönnberg T. A Practical Guide to Single-Cell Rna-Sequencing for Biomedical Research and Clinical Applications. *Genome Med* (2017) 9(1):1–12. doi: 10.1186/s13073-017-0467-4
51. Zhao S, Zhang Y, Gamini R, Zhang B, von Schack D. Evaluation of Two Main RNA-Seq Approaches for Gene Quantification in Clinical Rna Sequencing: Polya+ Selection Versus Rrna Depletion. *Sci Rep* (2018) 8 (1):1–12. doi: 10.1038/s41598-018-23226-4
52. Park J-E, Botting RA, Conde CD, Popescu D-M, Lavaert M, Kunz DJ, et al. A Cell Atlas of Human Thymic Development Defines T Cell Repertoire Formation. *Science* (2020) 367(6480). doi: 10.1101/2020.01.28.911115
53. Herzog VA, Reichholf B, Neumann T, Rescheneder P, Bhat P, Burkard TR, et al. Thiol-Linked Alkylation of RNA to Assess Expression Dynamics. *Nat Methods* (2017) 14(12):1198–204. doi: 10.1038/nmeth.4435

54. Erhard F, Baptista MA, Krammer T, Hennig T, Lange M, Arampatzi P, et al. Scslam-Seq Reveals Core Features of Transcription Dynamics in Single Cells. *Nature* (2019) 571(7765):419–23. doi: 10.1038/s41586-019-1369-y
55. Chen W, Zhao Y, Chen X, Yang Z, Xu X, Bi Y, et al. A Multicenter Study Benchmarking Single-Cell RNA Sequencing Technologies Using Reference Samples. *Nat Biotechnol* (2020) 1–12. doi: 10.1038/s41587-020-00748-9
56. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WMIII, et al. Comprehensive Integration of Single-Cell Data. *Cell* (2019) 177(7):1888–902.e21. doi: 10.1016/j.cell.2019.05.031
57. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, Sensitive and Accurate Integration of Single-Cell Data With Harmony. *Nat Methods* (2019) 16:1289–96. doi: 10.1038/s41592-019-0619-0
58. Liu T, Balzano-Nogueira L, Lleo A, Conesa A. Transcriptional Differences for COVID-19 Disease Map Genes Between Males and Females Indicate a Different Basal Immunophenotype Relevant to the Disease. *Genes* (2020) 11 (12):1447. doi: 10.3390/genes11121447
59. Wang X, Park J, Susztak K, Zhang NR, Li M. Bulk Tissue Cell Type Deconvolution With Multi-Subject Single-Cell Expression Reference. *Nat Commun* (2019) 10(1):1–9. doi: 10.1038/s41467-018-08023-x
60. Aguirre-Gamboa R, de Klein N, di Tommaso J, Claringbould A, van der Wijst MG, de Vries D, et al. Deconvolution of Bulk Blood Eqtl Effects Into Immune Cell Subpopulations. *BMC Bioinf* (2020) 21(1):1–23. doi: 10.1186/s12859-020-03576-5
61. Eraslan G, Simon LM, Mircea M, Mueller NS, Theis FJ. Single-Cell RNA-Seq Denoising Using a Deep Count Autoencoder. *Nat Commun* (2019) 10(1):1–14. doi: 10.1038/s41467-018-07931-2
62. Van Dijk D, Sharma R, Nainys J, Yim K, Kathail P, Carr AJ, et al. Recovering Gene Interactions From Single-Cell Data Using Data Diffusion. *Cell* (2018) 174(3):716–29. doi: 10.1016/j.cell.2018.05.061
63. Stoeckius M, Zheng S, Houck-Loomis B, Hao S, Yeung BZ, Mauck WM, et al. Cell Hashing With Barcoded Antibodies Enables Multiplexing and Doublet Detection for Single Cell Genomics. *Genome Biol* (2018) 19(1):1–12. doi: 10.1186/s13059-018-1603-1
64. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, et al. Large-Scale Simultaneous Measurement of Epitopes and Transcriptomes in Single Cells. *Nat Methods* (2017) 14(9):865. doi: 10.1038/nmeth.4380
65. Katzenelenbogen Y, Sheban F, Yalin A, Yofe I, Svetlichnyy D, Jaitin DA, et al. Coupled Scrna-Seq and Intracellular Protein Activity Reveal an Immunosuppressive Role of TREM2 in Cancer. *Cell* (2020) 182(4):872–85.e19. doi: 10.1016/j.cell.2020.06.032
66. Kumar NG, Contaifer D, Madurantakam P, Carbone S, Price ET, Van Tassell B, et al. Dietary Bioactive Fatty Acids as Modulators of Immune Function: Implications on Human Health. *Nutrients* (2019) 11(12):2974. doi: 10.3390/nu11122974
67. Loftus RM, Finlay DK. Immunometabolism: Cellular Metabolism Turns Immune Regulator. *J Biol Chem* (2016) 291(1):1–10. doi: 10.1074/jbc.R115.693903
68. Barba I, Fernandez-Montesinos R, Garcia-Dorado D, Pozo D. Alzheimer's Disease Beyond the Genomic Era: Nuclear Magnetic Resonance (Nmr) Spectroscopy-Based Metabolomics. *J Cell Mol Med* (2008) 12(5a):1477–85. doi: 10.1111/j.1582-4934.2008.00385.x
69. Dettmer K, Aronov PA, Hammock BD. Mass Spectrometry-Based Metabolomics. *Mass spectrometry Rev* (2007) 26(1):51–78. doi: 10.1002/mas.20108
70. Gorrochategui E, Jaumot J, Tauler R. Roimcr: A Powerful Analysis Strategy for LC-MS Metabolomic Datasets. *BMC Bioinf* (2019) 20(1):1–17. doi: 10.1186/s12859-019-2848-8
71. Fernández-Ochoa Á, Brunius C, Borrás-Linares I, Quirantes-Piné R, Cádiz-Gurrea M, Consortium PC, et al. Metabolic Disturbances in Urinary and Plasma Samples From Seven Different Systemic Autoimmune Diseases Detected by HPLC-ESI-QTOF-MS. *J Proteome Res* (2020) 19(8):3220–9. doi: 10.1021/acs.jproteome.0c00179
72. Kolmert J, Gómez C, Balgoma D, Sjödin M, Bood J, Konradsen JR, et al. Urinary Leukotriene E4 and Prostaglandin D2 Metabolites Increase in Adult and Childhood Severe Asthma Characterized by Type 2 Inflammation. A Clinical Observational Study. *Am J Respir Crit Care Med* (2021) 203(1):37–53. doi: 10.1164/rccm.202101-0208LE
73. Souter I, Bellavia A, Williams PL, Korevaar T, Meeker JD, Braun JM, et al. Urinary Concentrations of Phthalate Metabolite Mixtures in Relation to Serum Biomarkers of Thyroid Function and Autoimmunity Among Women From a Fertility Center. *Environ Health Perspect* (2020) 128(6):067007. doi: 10.1289/EHP6740
74. Bar N, Korem T, Weissbrod O, Zeevi D, Rothschild D, Leviatan S, et al. A Reference Map of Potential Determinants for the Human Serum Metabolome. *Nature* (2020) 588(7836):135–40. doi: 10.1038/s41586-020-2896-2
75. Al Bander Z, Nitert MD, Mousa A, Naderpoor N. The Gut Microbiota and Inflammation: An Overview. *Int J Environ Res Public Health* (2020) 17 (20):7618. doi: 10.3390/ijerph17207618
76. Fitzgibbon G, Mills KH. The Microbiota and Immune-Mediated Diseases: Opportunities for Therapeutic Intervention. *Eur J Immunol* (2020) 50 (3):326–37. doi: 10.1002/eji.201948322
77. Jiao Y, Wu L, Huntington ND, Zhang X. Crosstalk Between Gut Microbiota and Innate Immunity and its Implication in Autoimmune Diseases. *Front Immunol* (2020) 11:282. doi: 10.3389/fimmu.2020.00282
78. Zhang X, Li L, Butcher J, Stintzi A, Figeys D. Advancing Functional and Translational Microbiome Research Using Meta-Omics Approaches. *Microbiome* (2019) 7(1):1–12. doi: 10.1186/s40168-019-0767-6
79. Vujkovic-Cvijin I, Sklar J, Jiang L, Natarajan L, Knight R, Belkaid Y. Host Variables Confound Gut Microbiota Studies of Human Disease. *Nature* (2020) 587(7834):448–54. doi: 10.1038/s41586-020-2881-9
80. Frølund M, Wikström A, Lidbrink P, Abu Al-Soud W, Larsen N, Harder CB, et al. The Bacterial Microbiota in First-Void Urine From Men With and Without Idiopathic Urethritis. *PloS One* (2018) 13(7):e0201380. doi: 10.1371/journal.pone.0201380
81. Winters BR, Pleil JD, Angrish MM, Stiegel MA, Risby TH, Madden MC. Standardization of the Collection of Exhaled Breath Condensate and Exhaled Breath Aerosol Using a Feedback Regulated Sampling Device. *J breath Res* (2017) 11(4):047107. doi: 10.1088/1752-7163/aa8bbc
82. Cheung RK, Utz PJ. Cytof—the Next Generation of Cell Detection. *Nat Rev Rheumatol* (2011) 7(9):502–3. doi: 10.1038/nrrheum.2011.110
83. Pal A, Glaß H, Naumann M, Kreiter N, Japtok J, Sczech R, et al. High Content Organelle Trafficking Enables Disease State Profiling as Powerful Tool for Disease Modelling. *Sci Data* (2018) 5(1):1–15. doi: 10.1038/sdata.2018.241
84. Kiyoi T, Liu S, Sahid MNA, Shudou M, Maeyama K, Mogi M. High-Throughput Screening System for Dynamic Monitoring of Exocytotic Vesicle Trafficking in Mast Cells. *PloS One* (2018) 13(6):e0198785. doi: 10.1371/journal.pone.0198785
85. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the Multiple Testing Burden for Genomewide Association Studies of Nearly All Common Variants. *Genet Epidemiol* (2008) 32(4):381–5. doi: 10.1002/gepi.20303
86. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI Gwas Catalog of Published Genome-Wide Association Studies, Targeted Arrays and Summary Statistics 2019. *Nucleic Acids Res* (2019) 47(D1):D1005–D12. doi: 10.1093/nar/gky1120
87. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic Identification of Trans Eqtls as Putative Drivers of Known Disease Associations. *Nat Genet* (2013) 45(10):1238–43. doi: 10.1038/ng.2756
88. Consortium G. Genetic Effects on Gene Expression Across Human Tissues. *Nature* (2017) 550(7675):204–13. doi: 10.1038/nature24277
89. Min JL, Hemani G, Hannon E, Dekkers KF, Castillo-Fernandez J, Luijk R, et al. Genomic and Phenomic Insights From an Atlas of Genetic Effects on DNA Methylation. *medRxiv* (2020) 1:1–30. doi: 10.1101/2020.09.01.20180406
90. Xu C-J, Bonder MJ, Söderhäll C, Bustamante M, Baïz N, Gehring U, et al. The Emerging Landscape of Dynamic Dna Methylation in Early Childhood. *BMC Genomics* (2017) 18(1):1–11. doi: 10.1186/s12864-016-3452-1
91. Li Y, Oosting M, Deelen P, Ricaño-Ponce I, Smeekens S, Jaeger M, et al. Inter-Individual Variability and Genetic Influences on Cytokine Responses to Bacteria and Fungi. *Nat Med* (2016) 22(8):952–60. doi: 10.1038/nm.4139
92. Li Y, Oosting M, Smeekens SP, Jaeger M, Aguirre-Gamboa R, Le KT, et al. A Functional Genomics Approach to Understand Variation in Cytokine Production in Humans. *Cell* (2016) 167(4):1099–110.e14. doi: 10.1016/j.cell.2016.10.017
93. Hellwege JN, Keaton JM, Giri A, Gao X, Velez Edwards DR, Edwards TL. Population Stratification in Genetic Association Studies. *Curr Protoc Hum Genet* (2017) 95(1):1. doi: 10.1002/cphg.48
94. Martin ER, Tunc I, Liu Z, Slifer SH, Beecham AH, Beecham GW. Properties of Global-and Local-Ancestry Adjustments in Genetic Association Tests in Admixed Populations. *Genet Epidemiol* (2018) 42(2):214–29. doi: 10.1002/gepi.22103
95. Gamazon ER, Segrè AV, van de Bunt M, Wen X, Xi HS, Hormozdiari F, et al. Using an Atlas of Gene Regulation Across 44 Human Tissues to Inform

Complex Disease-and Trait-Associated Variation. *Nat Genet* (2018) 50 (7):956–67. doi: 10.1038/s41588-018-0154-4

96. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. A Gene-Based Association Method for Mapping Traits Using Reference Transcriptome Data. *Nat Genet* (2015) 47(9):1091. doi: 10.1038/ng.3367
97. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic Localization of Common Disease-Associated Variation in Regulatory Dna. *Science* (2012) 337(6099):1190–5. doi: 10.1126/science.1222794
98. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. The Encyclopedia of DNA Elements (Encode): Data Portal Update. *Nucleic Acids Res* (2018) 46(D1):D794–801. doi: 10.1093/nar/gkx1081
99. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative Analysis of 111 Reference Human Epigenomes. *Nature* (2015) 518(7539):317–30 doi: 10.1038/nature14248
100. Delaneau O, Zazhytska M, Borel C, Giannuzzi G, Rey G, Howald C, et al. Chromatin Three-Dimensional Interactions Mediate Genetic Effects on Gene Expression. *Science* (2019) 364(6439):eaat8266. doi: 10.1126/science.aat8266
101. McRae AF, Marioni RE, Shah S, Yang J, Powell JE, Harris SE, et al. Identification of 55,000 Replicated Dna Methylation Qtl. *Sci Rep* (2018) 8 (1):1–9. doi: 10.1038/s41598-018-35871-w
102. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic Atlas of the Human Plasma Proteome. *Nature* (2018) 558 (7708):73–9. doi: 10.1038/s41586-018-0175-2
103. Shin S-Y, Fauman EB, Petersen A-K, Krumsiek J, Santos R, Huang J, et al. An Atlas of Genetic Influences on Human Blood Metabolites. *Nat Genet* (2014) 46(6):543–50. doi: 10.1038/ng.2982
104. Nath AP, Ritchie SC, Byars SG, Fearnley LG, Havulinna AS, Joensuu A, et al. An Interaction Map of Circulating Metabolites, Immune Gene Networks, and Their Genetic Regulation. *Genome Biol* (2017) 18(1):1–15. doi: 10.1186/s13059-017-1279-y
105. Aguirre-Gamboa R, Joosten I, Urbano PC, van der Molen RG, van Rijssen E, van Cranenbroek B, et al. Differential Effects of Environmental and Genetic Factors on T and B Cell Immune Traits. *Cell Rep* (2016) 17(9):2474–87. doi: 10.1016/j.celrep.2016.10.053
106. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian Test for Colocalisation Between Pairs of Genetic Association Studies Using Summary Statistics. *PloS Genet* (2014) 10(5): e1004383. doi: 10.1371/journal.pgen.1004383
107. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An Atlas of Genetic Correlations Across Human Diseases and Traits. *Nat Genet* (2015) 47(11):1236. doi: 10.1038/ng.3406
108. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base Platform Supports Systematic Causal Inference Across the Human Phenome. *Elife* (2018) 7:e34408. doi: 10.7554/eLife.34408
109. Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martín D, et al. Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell* (2016) 167(5):1398–414. doi: 10.1016/j.cell.2016.10.026
110. Rosa M, Chignon A, Li Z, Boulanger M-C, Arsenault BJ, Bossé Y, et al. A Mendelian Randomization Study of IL6 Signaling in Cardiovascular Diseases, Immune-Related Disorders and Longevity. *NPJ genomic Med* (2019) 4(1):1–10. doi: 10.1038/s41525-019-0097-4
111. McGowan LM, Davey Smith G, Gaunt TR, Richardson TG. Integrating Mendelian Randomization and Multiple-Trait Colocalization to Uncover Cell-Specific Inflammatory Drivers of Autoimmune and Atopic Disease. *Hum Mol Genet* (2019) 28(19):3293–300. doi: 10.1093/hmg/ddz155
112. Baccarelli A, Bollati V. Epigenetics and Environmental Chemicals. *Curr Opin Pediatr* (2009) 21(2):243. doi: 10.1097/MOP.0b013e32832925cc
113. Martin DI, Cropley JE, Suter CM. Epigenetics in Disease: Leader or Follower? *Epigenetics* (2011) 6(7):843–8. doi: 10.4161/epi.6.7.16498
114. Ramos-Rodríguez M, Raurell-Vila H, Colli ML, Alvelos MI, Subirana-Granés M, Juan-Mateu J, et al. The Impact of Proinflammatory Cytokines on the β-Cell Regulatory Landscape Provides Insights Into the Genetics of Type 1 Diabetes. *Nat Genet* (2019) 51(11):1588–95. doi: 10.1101/560193
115. Netea MG, Joosten LA, Latz E, Mills KH, Natoli G, Stunnenberg HG, et al. Trained Immunity: A Program of Innate Immune Memory in Health and Disease. *Science* (2016) 352(6284):aaf1098. doi: 10.1126/science.aaf1098
116. Mazzone R, Zwergel C, Artico M, Taurone S, Ralli M, Greco A, et al. The Emerging Role of Epigenetics in Human Autoimmune Disorders. *Clin Epigenet* (2019) 11(1):1–15. doi: 10.1186/s13148-019-0632-2
117. Granja JM, Klemm S, McGinnis LM, Kathiria AS, Mezger A, Corces MR, et al. Single-Cell Multiomic Analysis Identifies Regulatory Programs in Mixed-Phenotype Acute Leukemia. *Nat Biotechnol* (2019) 37(12):1458–65. doi: 10.1038/s41587-019-0332-7
118. Ernst J, Kellis M. ChromHMM: Automating Chromatin-State Discovery and Characterization. *Nat Methods* (2012) 9(3):215–6. doi: 10.1038/nmeth.1906
119. Gjoneska E, Pfenning AR, Mathys H, Quon G, Kundaje A, Tsai L-H, et al. Conserved Epigenomic Signals in Mice and Humans Reveal Immune Basis of Alzheimer's Disease. *Nature* (2015) 518(7539):365–9. doi: 10.1038/nature14252
120. Cairns J, Freire-Pritchett P, Wingett SW, Várnai C, Dimond A, Plagnol V, et al. Chicago: Robust Detection of DNA Looping Interactions in Capture Hi-C Data. *Genome Biol* (2016) 17(1):1–17. doi: 10.1186/s13059-016-0992-2
121. Hu G, Cui K, Fang D, Hirose S, Wang X, Wangsa D, et al. Transformation of Accessible Chromatin and 3D Nucleome Underlies Lineage Commitment of Early T Cells. *Immunity* (2018) 48(2):227–42. doi: 10.1016/j.immuni.2018.01.013
122. Burren OS, García AR, Javierre B-M, Rainbow DB, Cairns J, Cooper NJ, et al. Chromosome Contacts in Activated T Cells Identify Autoimmune Disease Candidate Genes. *Genome Biol* (2017) 18(1):1–19. doi: 10.1186/s13059-017-1285-0
123. Chan WF, Coughlan HD, Zhou JH, Keenan CR, Bediaga NG, Hodgkin PD, et al. Pre-Mitotic Genome Re-Organisation Bookends the B Cell Differentiation Process. *Nat Commun* (2021) 12(1):1–13. doi: 10.1038/s41467-021-21536-2
124. Zhang J-Y, Wang X-M, Xing X, Xu Z, Zhang C, Song J-W, et al. Single-Cell Landscape of Immunological Responses in Patients With Covid-19. *Nat Immunol* (2020) 21(9):1107–18. doi: 10.1038/s41590-020-0762-x
125. Tian W, Zhang N, Jin R, Feng Y, Wang S, Gao S, et al. Immune Suppression in the Early Stage of COVID-19 Disease. *Nat Commun* (2020) 11(1):1–8. doi: 10.1038/s41467-020-19706-9
126. Wu TD, Madireddi S, de Almeida PE, Banchereau R, Chen Y-JJ, Chitre AS, et al. Peripheral T Cell Expansion Predicts Tumour Infiltration and Clinical Response. *Nature* (2020) 579(7798):274–8. doi: 10.1038/s41586-020-2056-8
127. Miller BC, Sen DR, Al Abosy R, Bi K, Virkud YV, LaFleur MW, et al. Subsets of Exhausted Cd8+ T Cells Differentially Mediate Tumor Control and Respond to Checkpoint Blockade. *Nat Immunol* (2019) 20(3):326–36. doi: 10.1038/s41590-019-0312-6
128. Setliff I, Shiakolas AR, Pilewski KA, Murji AA, Mapengo RE, Janowska K, et al. High-Throughput Mapping of B Cell Receptor Sequences to Antigen Specificity. *Cell* (2019) 179(7):1636–46.e15. doi: 10.1016/j.cell.2019.11.003
129. Nakaya HI, Hagan T, Duraisingham SS, Lee EK, Kwissa M, Rouphael N, et al. Systems Analysis of Immunity to Influenza Vaccination Across Multiple Years and in Diverse Populations Reveals Shared Molecular Signatures. *Immunity* (2015) 43(6):1186–98. doi: 10.1016/j.immuni.2015.11.012
130. Conesa A, Nueda MJ, Ferrer A, Talón M. Masigpro: A Method to Identify Significantly Differential Expression Profiles in Time-Course Microarray Experiments. *Bioinformatics* (2006) 22(9):1096–102. doi: 10.1093/bioinformatics/btl056
131. Bouhaddani S, Houwing-Duistermaat J, Salo P, Perola M, Jongbloed G, Uh HW.Evaluation of O2PLS in Omics Data Integration. *BMC Bioinf* (2016) 17: S11. doi: 10.1186/s12859-015-0854-z.
132. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The Dynamics and Regulators of Cell Fate Decisions are Revealed by Pseudotemporal Ordering of Single Cells. *Nat Biotechnol* (2014) 32(4):381. doi: 10.1038/nbt.2859
133. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed Graph Embedding Resolves Complex Single-Cell Trajectories. *Nat Methods* (2017) 14(10):979. doi: 10.1038/nmeth.4402
134. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. Rna Velocity of Single Cells. *Nature* (2018) 560(7719):494–8. doi: 10.1038/s41586-018-0414-6
135. Chua RL, Lukassen S, Trump S, Hennig BP, Wendisch D, Pott F, et al. Covid-19 Severity Correlates With Airway Epithelium–Immune Cell Interactions Identified by Single-Cell Analysis. *Nat Biotechnol* (2020) 38(8):970–9. doi: 10.1038/s41587-020-0602-4

136. Langfelder P, Horvath S. Wgcna: An R Package for Weighted Correlation Network Analysis. *BMC Bioinf* (2008) 9(1):559. doi: 10.1186/1471-2105-9-559

137. Deelen P, van Dam S, Herkert JC, Karjalainen JM, Brugge H, Abbott KM, et al. Improving the Diagnostic Yield of Exome-Sequencing by Predicting Gene–Phenotype Associations Using Large-Scale Gene Expression Analysis. *Nat Commun* (2019) 10(1):2837. doi: 10.1038/s41467-019-10649-4.

138. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, et al. Innatedb: Systems Biology of Innate Immunity and Beyond—Recent Updates and Continuing Curation. *Nucleic Acids Res* (2013) 41(D1): D1228–D33. doi: 10.1093/nar/gks1147

139. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING Database in 2017: Quality-Controlled Protein–Protein Association Networks, Made Broadly Accessible. *Nucleic Acids Res* (2017) 45(D1):D362–8. doi: 10.1093/nar/gkw937

140. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. Cellphonedb: Inferring Cell–Cell Communication From Combined Expression of Multi-Subunit Ligand–Receptor Complexes. *Nat Protoc* (2020) 15(4):1484–506. doi: 10.1038/s41596-020-0292-x

141. Corridoni D, Antanaviciute A, Gupta T, Fawkner-Corbett D, Aulicino A, Jagielowicz M, et al. Single-Cell Atlas of Colonic Cd8+ T Cells in Ulcerative Colitis. *Nat Med* (2020) 26(9):1480–90. doi: 10.1038/s41591-020-1003-4

142. Browaeys R, Saelens W, Saeys Y. Nichenet: Modeling Intercellular Communication by Linking Ligands to Target Genes. *Nat Methods* (2020) 17(2):159–62. doi: 10.1038/s41592-019-0667-5

143. Bonnardel J, T'Jonck W, Gaublomme D, Browaeys R, Scott CL, Martens L, et al. Stellate Cells, Hepatocytes, and Endothelial Cells Imprint the Kupffer Cell Identity on Monocytes Colonizing the Liver Macrophage Niche. *Immunity* (2019) 51(4):638–54. doi: 10.1016/j.immuni.2019.08.017

144. Cullen CM, Aneja KK, Beyhan S, Cho CE, Woloszynek S, Convertino M, et al. Emerging Priorities for Microbiome Research. *Front Microbiol* (2020) 11:136. doi: 10.3389/fmicb.2020.00136

145. Dorrestein PC, Mazmanian SK, Knight R. From Microbiomess to Metabolomes to Function During Host-Microbial Interactions. *Immunity* (2014) 40(6):824. doi: 10.1016/j.immuni.2014.05.015

146. Hattori M, Okuno Y, Goto S, Kanehisa M. Development of a Chemical Structure Comparison Method for Integrated Analysis of Chemical and Genomic Information in the Metabolic Pathways. *J Am Chem Soc* (2003) 125 (39):11853–65. doi: 10.1021/ja036030u

147. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, et al. Hmdb 4.0: The Human Metabolome Database for 2018. *Nucleic Acids Res* (2018) 46(D1):D608–17. doi: 10.1093/nar/gkx1089

148. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, et al. Metaboanalyst 4.0: Towards More Transparent and Integrative Metabolomics Analysis. *Nucleic Acids Res* (2018) 46(W1):W486–94. doi: 10.1093/nar/gky310

149. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, et al. Species-Level Functional Profiling of Metagenomes and Metatranscriptomes. *Nat Methods* (2018) 15(11):962–8. doi: 10.1038/s41592-018-0176-y

150. Schirmer M, Smeekens SP, Vlamakis H, Jaeger M, Oosting M, Franzosa EA, et al. Linking the Human Gut Microbiome to Inflammatory Cytokine Production Capacity. *Cell* (2016) 167(4):1125–36.e8. doi: 10.1016/j.cell.2016.10.020

151. Aden K, Rehman A, Waschina S, Pan W-H, Walker A, Lucio M, et al. Metabolic Functions of Gut Microbes Associate With Efficacy of Tumor Necrosis Factor Antagonists in Patients With Inflammatory Bowel Diseases. *Gastroenterology* (2019) 157(5):1279–92. doi: 10.1053/j.gastro.2019.07.025

152. Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, et al. Disease Variants Alter Transcription Factor Levels and Methylation of Their Binding Sites. *Nat Genet* (2017) 49(1):131–8. doi: 10.1038/ng.3721

153. Ananthakrishnan AN, Khalili H, Pan A, Higuchi LM, de Silva P, Richter JM, et al. Association Between Depressive Symptoms and Incidence of Crohn's Disease and Ulcerative Colitis: Results From the Nurses' Health Study. *Clin Gastroenterol Hepatol* (2013) 11(1):57–62. doi: 10.1016/j.cgh.2012.08.032

154. Bakker OB, Aguirre-Gamboa R, Sanna S, Oosting M, Smeekens SP, Jaeger M, et al. Integration of Multi-Omics Data and Deep Phenotyping Enables Prediction of Cytokine Responses. *Nat Immunol* (2018) 19(7):776–86. doi: 10.1038/s41590-018-0121-3

155. Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, et al. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* (2007) 129(4):823–37. doi: 10.1016/j.cell.2007.05.009

156. Marco A, Meharena HS, Dileep V, Raju RM, Davila-Velderrain J, Zhang AL, et al. Mapping the Epigenomic and Transcriptomic Interplay During Memory Formation and Recall in the Hippocampal Engram Ensemble. *Nat Neurosci* (2020) 23:1606–17. doi: 10.1038/s41593-020-00717-0

157. Hernández-de-Diego R, Tarazona S, Martínez-Mira C, Balzano-Nogueira L, Furió-Tarí P, Pappas GJJr., et al. Paintomics 3: A Web Resource for the Pathway Analysis and Visualization of Multi-Omics Data. *Nucleic Acids Res* (2018) 46(W1):W503–9. doi: 10.1093/nar/gky466

158. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets. *Nat Commun* (2019) 10(1):1–10. doi: 10.1038/s41467-019-09234-6

159. Akhmedov M, Martinelli A, Geiger R, Kwee I. Omics Playground: A Comprehensive Self-Service Platform for Visualization, Analytics and Exploration of Big Omics Data. *NAR Genomics Bioinf* (2020) 2(1):lqz019. doi: 10.1093/nargab/lqz019

160. Aguilar SV, Aguilar O, Allan R, Amir EAD, Angeli V, Artyomov MN, et al. Immgen at 15. *Nat Immunol* (2020) 21(7):700–3. doi: 10.1038/s41590-020-0687-4

161. Scholz CJ, Biernat P, Becker M, Baßler K, Günther P, Balfer J, et al. Fastgenomics: An Analytical Ecosystem for Single-Cell RNA Sequencing Data. *bioRxiv* (2018) 1:272476. doi: 10.1101/272476

162. Szymczak S, Dose J, Torres GG, Heinsen F-A, Venkatesh G, Datlinger P, et al. Dna Methylation Qtl Analysis Identifies New Regulators of Human Longevity. *Hum Mol Genet* (2020) 29(7):1154–67. doi: 10.1093/hmg/ddaa033

163. Huan T, Joehanes R, Song C, Peng F, Guo Y, Mendelson M, et al. Genome-Wide Identification of DNA Methylation Qtls in Whole Blood Highlights Pathways for Cardiovascular Disease. *Nat Commun* (2019) 10(1):1–14. doi: 10.1038/s41467-019-12228-z

164. Morrow J, Qiu W, Make B, Regan E, Han M, Hersh C, et al. DNA Methylation Is Predictive of Mortality in Current and Former Smokers. *Am J Respir Crit Care Med* (2020) 201(9):1099–109. doi: 10.1164/rccm.201902-0439OC

165. Carrier M, Robert M-È, González Ibáñez F, Desjardins M, Tremblay M-È. Imaging the Neuroimmune Dynamics Across Space and Time. *Front Neurosci* (2020) 14:903. doi: 10.3389/fnins.2020.00903

166. Chu C, Artis D, Chiu IM. Neuro-Immune Interactions in the Tissues. *Immunity* (2020) 52(3):464–74. doi: 10.1016/j.immuni.2020.02.017

167. Stakenborg N, Viola MF, Boeckxstaens GE. Intestinal Neuro-Immune Interactions: Focus on Macrophages, Mast Cells and Innate Lymphoid Cells. *Curr Opin Neurobiol* (2020) 62:68–75. doi: 10.1016/j.conb.2019.11.020

168. Delhalle S, Bode SF, Balling R, Ollert M, He FQ. A Roadmap Towards Personalized Immunology. *NPJ Syst Biol Appl* (2018) 4(1):1–14. doi: 10.1038/s41540-017-0045-9

169. Karakike E, Giamarellos-Bourboulis EJ. Macrophage Activation-Like Syndrome: A Distinct Entity Leading to Early Death in Sepsis. *Front Immunol* (2019) 10:55. doi: 10.3389/fimmu.2019.00055

170. Donovan J, Phu NH, Le Thi Phuong Thao NH, Lan NTHM, Trang NTM, Hiep NTT, et al. Adjunctive Dexamethasone for the Treatment of HIV-Uninfected Adults With Tuberculous Meningitis Stratified by Leukotriene A4 Hydrolase Genotype (Last ACT): Study Protocol for a Randomised Double Blind Placebo Controlled non-Inferiority Trial. *Wellcome Open Res* (2018) 3:32. doi: 10.12688/wellcomeopenres.14007.1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

REVIEW
published: 31 August 2021
doi: 10.3389/fimmu.2021.714461

Application of Genetic Studies to Flow Cytometry Data and Its Impact on Therapeutic Intervention for Autoimmune Disease

Valeria Orrù[1]*, Maristella Steri[1], Francesco Cucca[1,2] and Edoardo Fiorillo[1]

[1] Institute for Genetic and Biomedical Research, National Research Council (CNR), Sardinia, Italy, [2] Department of Biomedical Sciences, University of Sassari, Sassari, Italy

In recent years, systematic genome-wide association studies of quantitative immune cell traits, represented by circulating levels of cell subtypes established by flow cytometry, have revealed numerous association signals, a large fraction of which overlap perfectly with genetic signals associated with autoimmune diseases. By identifying further overlaps with association signals influencing gene expression and cell surface protein levels, it has also been possible, in several cases, to identify causal genes and infer candidate proteins affecting immune cell traits linked to autoimmune disease risk. Overall, these results provide a more detailed picture of how genetic variation affects the human immune system and autoimmune disease risk. They also highlight druggable proteins in the pathogenesis of autoimmune diseases; predict the efficacy and side effects of existing therapies; provide new indications for use for some of them; and optimize the research and development of new, more effective and safer treatments for autoimmune diseases. Here we review the genetic-driven approach that couples systematic multi-parametric flow cytometry with high-resolution genetics and transcriptomics to identify endophenotypes of autoimmune diseases for the development of new therapies.

Keywords: drug development, immune profiling, autoimmune diseases, GWAS, flow cytometry

OPEN ACCESS

Edited by:
Lam Cheung Tsoi,
University of Michigan, United States

Reviewed by:
Christoph Hudemann,
Philipps-University Marburg, Germany
Leonid Padyukov,
Karolinska Institutet (KI), Sweden
Francesco Annunziato,
University of Florence, Italy

****Correspondence:***
Valeria Orrù
valeria.orru@irgb.cnr.it

Specialty section:
This article was submitted to Autoimmune and Autoinflammatory Disorders, a section of the journal Frontiers in Immunology

Received: *25 May 2021*
Accepted: *13 August 2021*
Published: *31 August 2021*

Citation:
Orrù V, Steri M, Cucca F and Fiorillo E (2021) Application of Genetic Studies to Flow Cytometry Data and Its Impact on Therapeutic Intervention for Autoimmune Disease. Front. Immunol. 12:714461. doi: 10.3389/fimmu.2021.714461

INTRODUCTION

The human immune system is a magnificent biological network of specialized cells and their soluble products that can recognize and tolerate "self" and harmless symbionts while mounting responses to "non-self", including the panoply of harmful pathogens. Immune cell subtypes are the pivotal determinant to maintain immunity and minimize the loss of tolerance that can result in autoimmunity. Because immune cells must orchestrate and mount responses to a variety of insults, their circulating levels are extensively regulated by exposure to environmental factors, and in particular by pathogen infection. Nevertheless, in the last 10 years the assessment of genetic effects on circulating levels of immune cells and their surface proteins (collectively referred as immune cell traits) has revealed that they are on average ~40% heritable (1, 2), meaning that a high percentage of variability in their levels is regulated by genetic differences among individuals. The high heritability of immune cell traits has prompted us and others (1–6) to assess the genetic contribution to their variability through systematic genome wide association studies (GWAS) in general populations. Overall, hundreds of associated variants have been identified.

More recently, a GWAS-based approach on cytometric data has also been applied, albeit in a small sample size, to assess the genetic control of changes in immune cell levels after exposures such as influenza vaccination (7). This type of analysis is likely to become increasingly common and performed in much larger sample sets, for example to assess cellular response to the Sars-Cov-2 vaccine. There are three key requirements to use the powerful and unbiased tool of GWAS to understand how the immune cells are genetically regulated and to identify overlaps with autoimmune disease risk. The first is a very detailed measurement of a broad spectrum of cell types, encompassing innate and adaptive immunity, by assessing their activated, regulatory, inflammatory and maturation states. The second is high-resolution characterization of genetic variability in the same individuals. The third requirement is generating or obtaining summary statistic data of autoimmune disease GWAS to establish overlap with immune cell GWAS. The sample size of immune cell GWAS is pivotal to infer a full range of genetic associations. Indeed, while a few thousand individuals, like those assessed in the immune cell trait GWAS performed thus far, identify genetic associations of common variants with relatively large effect size, tens of thousands of individuals must be analyzed to discover genetic associations with rare variants, and those with smaller effect size (8). Further broadening the spectrum of associated variants through substantial increases in the sample size evaluated in immune cell trait GWAS will thus be important to identify many more overlapping associations with disease. Of particular interest are multiple overlaps with the same immune trait and disease, strengthening the evidence for a causal relationship and thereby increasing the power to identify therapeutic targets.

Focusing on immune cell traits, the most common technique to systematically measure cell subpopulations as well as surface or intracellular proteins, is flow cytometry. Routinely used for functional studies, flow cytometry is now becoming the starting point to identify DNA variants associated with immune traits and, in turn, those variants that are also associated with risk of disease (hereafter referred to as "overlapping genetic associations"). This approach can identify cell types, molecules and pathways implicated in disease pathogenesis and provide prime candidates for more specific and efficacious therapeutic intervention (**Figure 1**). The potential of the genetic-driven approach in the research and development of new drugs is supported by the observation that 73% of studies supported by genetic evidence targeting the disease pathway were successful in Phase II clinical trials compared with 43% of studies without such genetic link (9). Nevertheless, genetic studies provide only a powerful substrate for experimental elucidation of disease mechanisms. Thus, causality must be confirmed by functional experiments *in vitro* and *in vivo*, which, in the context discussed here, are essential to clarify the biological mechanisms underlying the overlapping associations with specific immune cell traits and disease risk and formulate robust therapeutic hypotheses that are critical to the success of new drug research and development programs.

In particular, genetic associations of quantitative cellular traits and autoimmune diseases are more likely to give rise to biological

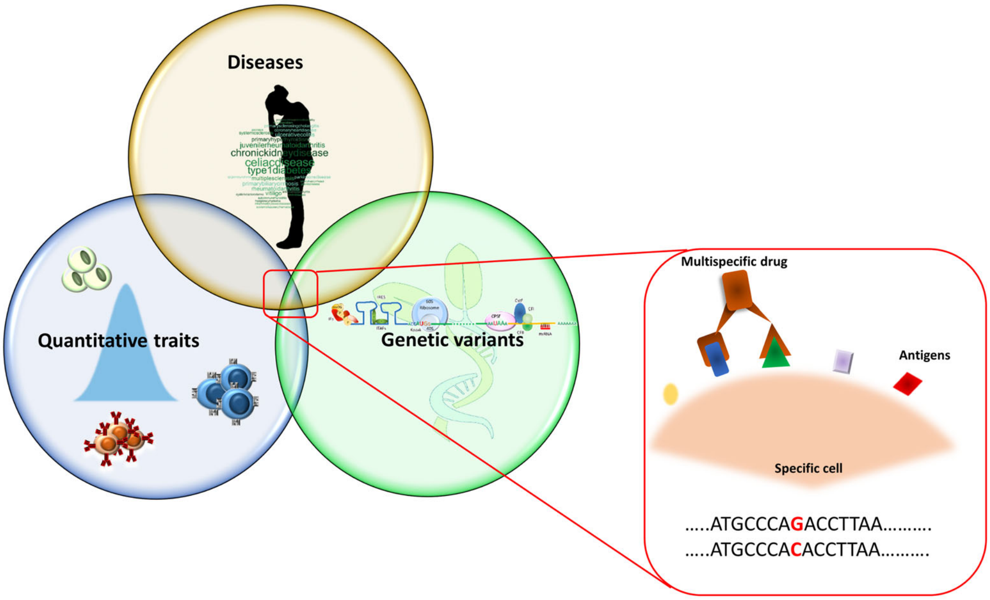


FIGURE 1 | Overview of the study approach. The picture summarizes the research of overlapping genetic associations that initially consists in the identification of genetic signals regulating immune cells in the general population. These signals are then compared with those associated with diseases with the aim to identify the genetic variants both affecting immune cells and disease risk. The identified immune cells can be considered as potential drug target on which to act pharmacologically. Being more frequently detectable by more than one antigen expressed simultaneously on their surface, these cells can be targeted by multi-specific drugs (binding two or more specific antigens).

investigations that are truly related to the causal biology of diseases than epidemiological surveys of environmental factors and observational studies of phenotypic variables that can often highlight second-order phenomena that are a consequence, and not a cause, of the disease. In this sense, although epidemiological evidence clearly indicates that environmental factors should play a very important role in the regulation of the immune system and contribute to the risk of autoimmune diseases, their precise identification is complicated by numerous factors and remains largely elusive. In contrast, genetics represents a more direct, powerful, and unbiased tool to generate robust hypotheses about disease-causing mechanisms that need to be further investigated with functional studies to identify and validate therapeutic targets (10).

We turn to an outline of the evolution of flow cytometry; the proper generation of flow cytometry data; and the application of GWAS to flow cytometry-based immune profiling to identify new drug targets.

FLOW CYTOMETRY

The role of flow cytometry (**Figure 2**) in scientific research and clinical practice is increasing dramatically and only a marginal part of its potential is currently being used. However, while this technique is very useful if applied correctly and with appropriate checks, it can lead to incorrect conclusions if not. We will dedicate the next two sub-sections to describe this technology and some tips for using it properly.

Flow Cytometry From Its Inception to Today

Flow cytometry development (11) was accompanied by important evolution of its applications in several scientific fields, including not only immunology but also hematology, cancer, microbiology, and physics. For instance, flow cytometric oncology panels are widely used to diagnose hematologic malignancy, especially B cell lymphoproliferative disorders, based on disproportion of kappa and lambda immunoglobulin light chains that are expressed on membrane surface of B cells. Indeed, a kappa-lambda ratio higher than 3:1 or lower than 1:3 is respectively considered evidence of monoclonality and diagnostic for B cell lymphoproliferative disorders (12).

In microbiology, flow cytometry allows the detection of microbes, their viability and distribution within cells that can have profound impact in infection diagnosis (13). Furthermore, in some countries, application of flow cytometry to microbiology has been routinely applied to water quality analysis (14).

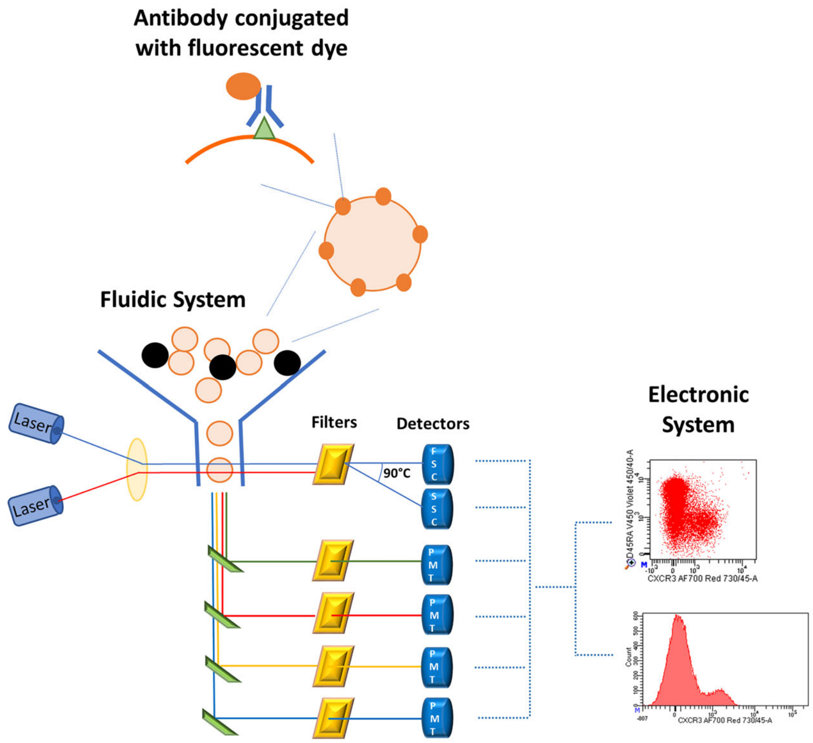


FIGURE 2 | Schematic representation of flow cytometry system. Cells stained with fluorescent-conjugated antibodies are aligned in the fluidic system where they encounter one or more laser beams which excite the fluorescent dyes bound to the cells. The fluorescent antibodies emit at a specific wavelength and the emission is proportional to the amount of antigen-antibody complex. The emission arrives to the optical system, consisting of filters, mirrors, and photomultiplier tubes (PMTs), which enhance and improve the signal. Finally, the electronic system converts the fluorescent emission in electronic signals visualized by histograms or bi-dimensional plots.

Improving performance and processivity and increasing the number of parameters measured simultaneously by flow cytometers is the major challenge for flow cytometry companies. For instance, to reduce the time of sample processing and the variability of data acquisition, an acoustic focusing chamber characterized by high frequency sound produced by a piezoelectric device were applied to a flow cytometer (15). This system generates a standing wave in the sample capillary, which can align cells in the center of the flux even when the original cell concentration is high.

To increase the number of antibodies assessed simultaneously, an alternative cytometry-based technique, namely "CyTOF" (cytometry by time-of-flight), was developed about ten years ago. Similarly, to flow cytometry, antigens are recognized by antibodies labeled with heavy metal isotopes (instead of fluorochromes) which, as in mass spectrometry, are detected based of on their time-of-flight (16). CyTOF is more expensive than classical flow-cytometry, require longer period of time to process each sample, making this method unsuitable for processing large amounts of samples in a short time, but it can detect more than 100 parameters per cell simultaneously.

Flow cytometry has also become the starting point for big data projects such as genetic studies of thousands of immune cell traits, and single cell transcriptomic and proteomic measurements. Moreover, the simultaneous assessment of several fluorochrome-conjugated antibodies (17) (destined to increase soon) in thousands of individuals allows the identification of very rare cell subsets and of new cell types never previously described, but at the same time, it increases the difficulty of analysis of the enormous amount of data generated. Indeed, to visualize an n-dimensional flow data, $^1/_2 \times n \times (n - 1)$ bi-dimensional plots would be needed, so that, for instance, an experiment assessing 20 antibodies would require $^1/_2 \times 20 \times (20 - 1) = 190$ bi-dimensional plots to display all marker combinations. Thus, data produced by the latest generation flow cytometry and CyTOF need to be visualized in alternative ways, departing from the classical bi-dimensional plots and histograms (**Figures 3A, B**).

Two of the most popular algorithms to reduce the complexity of this big amount of data and to identify populations of interest are SPADE (spanning-tree progression analysis of density-normalized events) (18) and t-SNE (t-stochastic neighbor embedding) (19). Both resolve high-dimensional data into a single bi-dimensional plot, the former visualizing cell clusters through dendrograms and the latter by scatter plots, so that the closer the cell clusters are, the more similar they are (**Figures 4A, B**).

SPADE and t-SNE do not allocate every cell to a specific cluster, nevertheless, automated clustering algorithms such as ACCENSE (20), DensVM (21), viSNE (22), to mention only a few of them, can help to solve this issue. However, these algorithms do not consider the entire dimension of the dataset; to address this, PhenoGraph was developed (**Figure 4C**) (23).

Another algorithms, named Wanderlust (24) is particularly useful to study temporal developmental cell relationships by

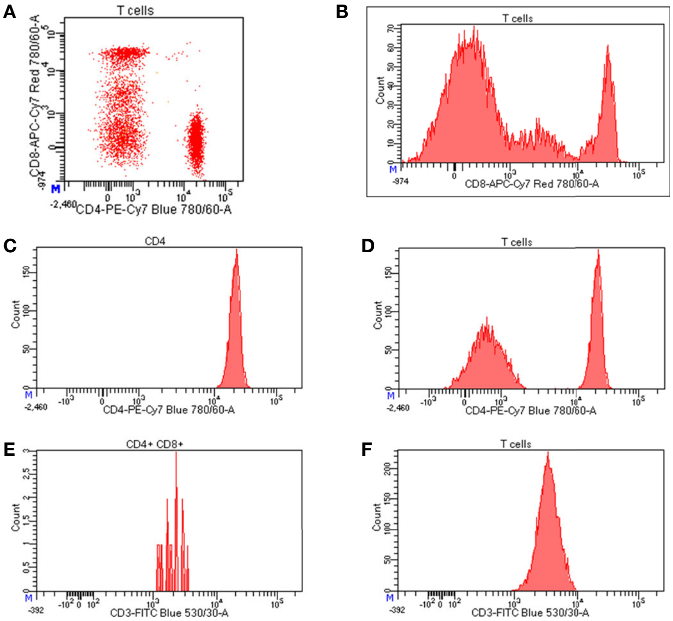


FIGURE 3 | Representation of flow cytometry data. **(A)** Bi-dimensional visualization of data (dot plot) where each axis represents an antigen; **(B)** histograms representing the expression level of CD8 on T cells; starting from left to right, the first peak corresponds to CD8 negative T cells, the second peak represents cells expressing intermediate level of CD8, whereas the third peaks indicates highly positive cells for CD8 expression; **(C)** normal distribution of CD4 expression on CD4 positive cells; **(D)** bimodal distribution of CD4 expression on T cells where the peak on the left corresponds to CD4 negative T cells, while the peak on the right represents CD4 positive T cells; expression levels of CD3 on **(E)** a poorly represented cell population (CD4+CD8+ T cells) and **(F)** a well-represented cell population (T cells).

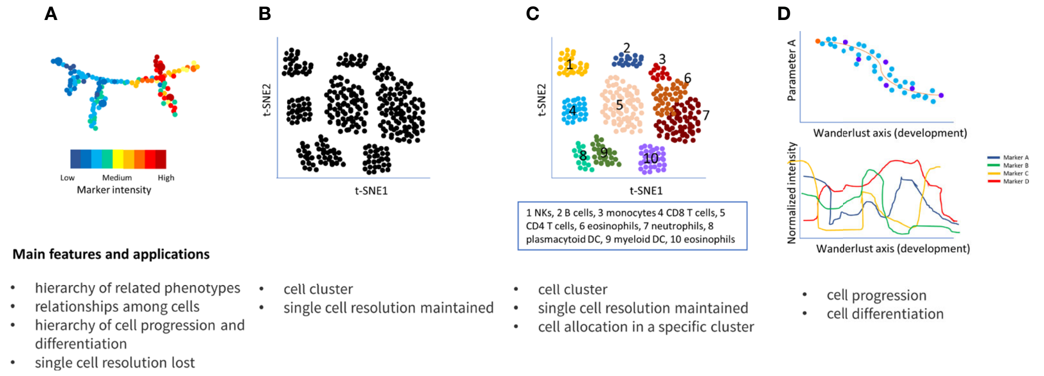


FIGURE 4 | Main approaches to resolve flow cytometry data complexity. **(A)** SPADE connects clusters of multidimensional data in a progressive dendrogram. Cluster sizes correlate with the number of cells within the cluster. The heat map indicates the intensity of each cluster based on the median intensities of a protein marker in each cell node; **(B)** t-SNE detects cluster corresponding to cell population, similar cell are placed close together reflecting their proximity in high-dimensional space; **(C)** vi-SNE, ACCENSE, DensVM and Phenograph are evolution of t-SNE and similarly visualized; in particular, Phenograph is able to assign each cell into a specific cluster; **(D)** Wanderlust orders cells into a trajectory corresponding to their developmental stages.

generating a trajectory, for example ranging from hematopoietic stem cell through the mature status of the assessed cells (**Figure 4D**). Both PhenoGraph and Wanderlust represent each cell by a node that is linked to its neighbors by edges; thus, phenotypically similar cell clusters are visualized by interconnected nodes, namely “neighborhoods” or “communities” of cells (25).

In case of comparisons among two or more groups (such as patients and controls), Citrus is another useful tool to identify differential cell clusters and response features among the assessed groups that could be predictive of different experimental or clinical endpoints of interest (26). For instance, comparing unstimulated *vs* stimulated peripheral blood mononuclear cells, Citrus was able to identify 117 cluster features (out of 465) which differed between the two conditions.

Guides to Correct Flow Cytometry Analysis

Before starting data collection and analysis, a strict process of quality checks and controls is pivotal to obtain reproducible and robust results. The most important steps can be summarized as follows.

1) *Panel set-up*. Increasingly, a number of common antigens are found to be expressed in cells whose biological role is supposed to be radically different. For instance, Schuh and colleagues described the uncommon co-expression of CD3 (receptor complex characterizing T cells) and CD20 (characterizing B cells) in a small subset of circulating lymphocytes that are especially frequent in the cerebrospinal fluid of multiple sclerosis patients (27). This underlines the need for several cell antigens simultaneously assessed as mandatory for a comprehensive immune cell analysis and for the discovery of rare cell populations that may nevertheless be potentially relevant in disease predisposition. However, the simultaneous assessment of many antigens requires a complex panel set-up that implies careful selection of antigen-fluorochrome combinations. A general role for fluorochrome-antigen selection is to use weak fluorochromes for highly expressed antigens and, vice-versa, bright fluorochromes for weakly expressed antigens. This allows detection of weak signals while keeping on scale brighter ones and minimizing the spillover of one fluorochrome into those having close emission wavelength. The mathematical correction of this spillover is called compensation and is an extremely important step that must be done before analyzing data to avoid misleading interpretations (28).
2) *Processing of samples*. The protocol to be followed and the time between sample collection and processing are pivotal to ensure reproducibility of flow data, especially for specific cells and antigens. For instance, monocytes are prone to modify their morphology and the expression of some antigens on their surface, including the costimulatory molecules CD80 and CD86 (29), while platelets are subject to very fast modifications and activation. Thus, this blood component should be processed within minutes after blood collection (30, 31). Similarly, the stability of antibodies is important: the Lyotube™ technology, employing lyophilized predefined cocktails of antibodies, is more stable than corresponding liquid formats, thus minimizing fluorochrome decay and allowing reduction of potential operator-dependent variations (1, 32).
3) *Sample freezing*. Freezing is known to damage some antigens and cell types, such as myeloid derived suppressor cells (defined as CD66b+ and CD15+, HLA-DRdim and CD14–) that are not detectable in previously frozen peripheral blood mononuclear cells (33). Special care should be taken to compare fresh with frozen samples, and as good practice it is strongly recommended to perform preliminary experiments to verify the quality/status of each antigen of interest before and after freezing.
4) *Systematic controls to monitor analyzer performance*. Flow cytometers are subject to laser wear and fluidic instability over time.

To compare samples acquired in different days, several controls should be used to ensure the correct and constant performance of flow cytometer and the consistency of data collection. Indeed, some analyzers are equipped with a system that performs daily electronic checks and automatically adjusts internal parameters.

Furthermore, reference stabilized blood samples with defined ranges of the main lymphocyte subsets are available to be used as controls, helping to avoid batch effects.

Once the samples have been acquired and processed in the proper way, the next step is gating. There are several ways to gate samples:

-Manual
-Semi-automatic
-Automatic

Each method has advantages and disadvantages; for instance, if, on the one hand manual gating is time-consuming and operator-dependent, on the other it allows the analysis of very rare cell populations that are difficult to identify using automatic strategies. Automatic methods (briefly described in the previous section) use algorithms to systematically identify cell populations, thus avoiding operator-dependent inaccuracies. They can be further divided into "hypothesis dependent", if the scientist sets specific cell subtypes to be measured, and "agnostic", which are not based on specific hypotheses, allowing the identification of previously unknown cell cluster which could be missed by using manual gating approaches.

Following gate positioning, each cell population (both newly identified and already known) can undergo three types of measurements:

a) Relative count
b) Absolute count
c) Fluorescence intensity

a) The relative count corresponds to the ratio between cell types that could be hierarchically dependent (e.g., percentage with respect to parental and grand parental cell population, such as percentage of CD4 with respect to T cells) or independent (e.g., ratio between T and B cells).

b) The absolute or actual count corresponds to the number of cells per volume (generally expressed as cells/ul or cell/mm^3). In human blood, the necessary condition to obtain absolute counts is to process fresh non-washed samples and use either analyzers able to calculate the absolute number of cells based on sample volume or a fixed number of counting beads to be added to each sample. In the latter case, it is necessary to apply a simple proportion between number of beads and cells acquired to obtain actual counts. Alternatively, it is also possible to obtained actual counts from frozen samples if the leukocyte (or lymphocyte) count measured on the day of the withdrawal is combined with the relative counts obtained by flow cytometry from frozen material.

c) Generally defined as mean or median fluorescence intensity (MFI), it represents the expression level of an antigen (such as CD4, CD8, CD40, CD28) on a cell type (**Figure 3C**). A necessary condition to properly analyze MFIs is that the marker measurement in the specific cells follows a normal (Gaussian) distribution. For instance, CD4 expression measured in total T cells (which include an important amount of CD4 negative cells) is inaccurate because a bimodal distribution would be observed: one peak corresponding to CD4 negative cells and a second peak corresponding to positive cells (**Figure 3D**). In this case, the bimodal distribution does not mirror the expression level of CD4 positive cells; rather, it correlates the number of cells present in the first (negative) peak with respect to the second (positive) one (**Figure 3D**). Thus, CD4 MFI should be assessed only in the CD4 positive cells, where its distribution is normal. In additional cases, such as CD8 expression in T cells, the presence of three peaks is frequently observed, corresponding to negative, intermediate (dim), and high (bright) antigen expression. The negative peak should be excluded, while the expression of CD8 in the two positive peaks should be measured separately, especially if the number of CD8 dim T cells is consistently represented (**Figure 3B**). Also, the number of events in which the MFI is measured is very important to obtain reliable data, as the MFI of a few events is not very robust. Thus, also in this case, the general rule is that the more events acquired, the more robust the MFI data are (**Figures 3E, F**).

UNDERSTANDING CAUSAL EFFECTS OF IMMUNE CELL LEVELS IN HUMAN DISEASE: THE HYPOTHESIS-GENERATING *VS* HYPOTHESIS-DRIVEN APPROACH

The comparison of specific immune cell levels between cases and controls has been a widely used approach to identify those cells or derived parameters that are more frequent in cases, and thus putatively predisposing to the disease, compared to controls. By contrast, those that are higher in controls are putatively protective for the disease. However, this case-control, hypothesis-driven comparison of immune phenotypes is limited by *a priori* knowledge and is also affected by second order effects due to the disease process and the administered therapy. That can lead to mistaken inference of a consequence of a disease for a cause (so-called *reverse causation*).

A more robust and systematic approach to identify immune cell traits implicated in the disease process relies on correlations between genetic association signals detected in different sample sets. This hypothesis-generating approach first establishes, *via* quantitative trait locus (QTL) GWAS, the genetic control of as many immune cell traits as possible in as many general population individuals as possible. The resulting association signals for immune cell traits are then evaluated for any significant overlap with association signals from GWAS on autoimmune disease risk, typically performed with a case-control design. The increasing availability of autoimmune disease GWAS summary statistics offers valuable resource data to search for such overlaps, which can then be formally demonstrated using specific statistical approaches like *co-localization* methods. These allow to formally test whether two association signals at the same locus for two different traits or diseases share the same causal variant (34). In principle, if a gene variant X is causally related to both a quantitative immunophenotype Y and an autoimmune disease Z, it is possible that the immunophenotype Y is involved in the process leading to the autoimmune disease Z and represents an endophenotype for that disease.

The route toward unequivocally linking a given immune cell variable with one or more immune mediated disease is rather complex and hindered by several factors including pleiotropic effects, low statistical power and incomplete characterization of immune cell variation, as follows (**Figure 5**).

Pleiotropy (**Figure 5A**), a phenomenon in which one genetic locus influences two or more phenotypic traits (35), is an emerging feature of current GWAS results that can complicate the resolution of the causal-relationships to a true disease-related intermediate immune phenotype (3). It is classically divided into biological or mediated, with the former referring to a genetic variant that has a direct influence on the regulation of more than one trait and the latter occurring when a variant directly influences one trait, which in turn influences another trait. Pleiotropy can also be spurious, which is due to various design artifacts that cause a genetic variant to appear fallaciously associated with multiple traits.

During the last decade, 93 loci associated with immune cells traits have been identified by genome wide association studies (4, 5, 36–38), and about half of these loci overlap with previously reported disease-associations predominantly for autoimmune disorders. Most of the detected genetic signals were characterized by pleiotropy; 61% of these signals regulate protein levels on the cell membrane (MFIs), whereas only 25% and 14% of them were found associated with relative and absolute counts, respectively (3). This can likely occur either because of the common origin and shared mechanisms of genetic regulation of different immune cells or because of the interrelated functions of many immune cell types, with some cells controlling the level of other cells. And the complexity of genetic associations detected so far with the genetic regulation of immune traits goes beyond the detection of pleiotropic effects and includes several instances of multiple independent signals in a given gene region affecting the same cell or protein expression, and in other cases unrelated traits (**Figure 5B**, also see the CD25 example in the next section).

In the presence of strong pleiotropy, approaches that exploit Mendel's second law of inheritance to search for *multiple independent genetic associations* associated with both the same intermediate immune phenotype and autoimmune disease outcome provide a route to somewhat restrict the number of coincident associations to those most likely involved in disease pathogenesis. Indeed, if two or more independent genetic signals are simultaneously associated with the same disease predisposition and a specific quantitative trait, with a coherent reduction or increase in the trait levels, it is more probable that the trait is causally implicated

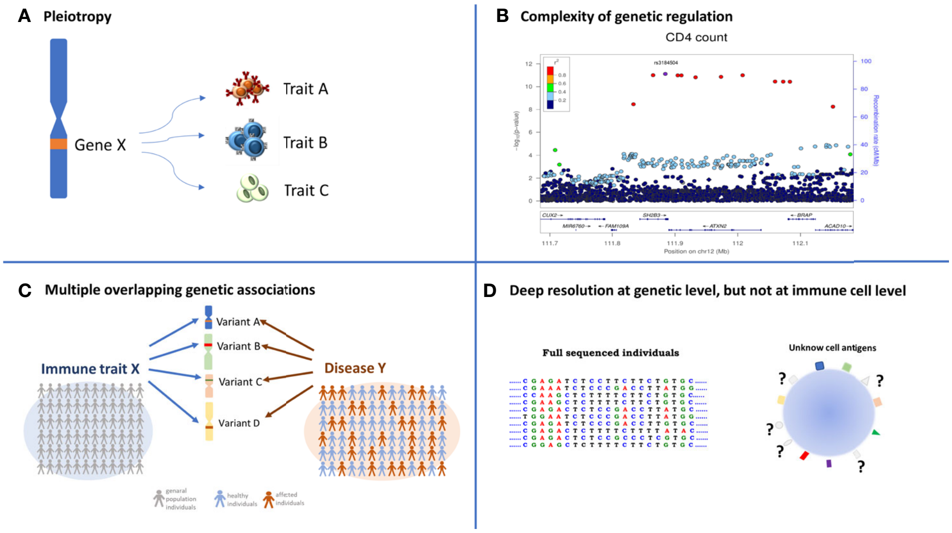

A Pleiotropy
Gene X
Trait A
Trait B
Trait C
B Complexity of genetic regulation
CD4 count
r²
0.8
0.6
0.4
0.2
rs3184504
−log10(p-value)
0
2
4
6
8
10
12
Recombination rate (cM/Mb)
0
20
40
60
80
100
CUX2→
MIR6760→
←FAM109A
SH2B3→
←ATXN2
←BRAP
ACAD10→
111.7
111.8
111.9
112
112.1
Position on chr12 (Mb)
C Multiple overlapping genetic associations
Immune trait X
Variant A
Variant B
Variant C
Variant D
Disease Y
genaral population individuals
healthy individuals
affected individuals
D Deep resolution at genetic level, but not at immune cell level
Full sequenced individuals
......C G A G A T C T C C T T C T T C T G T G C......
......C G A A A T C T C C C G A C C T T A T G G......
......C C A A G C T C T T T T C T T C T G T G C......
......C G A A G C T C T T T T C T T C T G T G C......
......C G A G A C T C T C C G A C C T T A T G C......
......T G G A A T C T C C C G A C C T T A T G G......
......C G A G A T C T C C C G A C C T T G T G C......
......C G A G A C T C T T T T C T T T T A T A C......
......C G A G A C T C T C C G C C C T C G T G C......
......C G G A G C T C T T T T C T T C T G T G C......
Unknow cell antigens
?
?
?
?
?

FIGURE 5 | Complexity in identifying overlapping association between immune trait and diseases. The four quadrants summarize the layer of difficulties that must be considered when the overlapping association approach is applied. **(A)** Pleiotropy: a specific variant can regulate several traits; **(B)** the identification of the causal variant is often not immediate due to the high LD among variants. Each dot represents a genetic variant, LD among variants (expressed as r^2) is color coded and specified in the legend, the significance of the association, expressed as –log10[P value], is indicated in the left y axis, and the genomic positions in the x axis; **(C)** the identification of several independent genetic variants associated with the same immune cell X and the same disease Y (multiple overlapping genetic association) requires thousands of deeply immune profiled individuals, both from general populations to dissect immune cell genetic regulation, and from case-control studies to identify genetic regulation of diseases; **(D)** the genomic information of each individual is very deep being at single-base resolution, whereas the knowledge of all the proteins expressed on cell surface is far from being complete.

in the disease predisposition (**Figure 5C**) (1, 3). This approach can help identify the most promising association signals to follow up with downstream functional studies; however, it does not reveal the presence of confounding factors regulating both trait and disease, because it is based on a simple comparison among a few association statistics. For these reasons, *Mendelian randomization* (MR) is now the approach most frequently applied to infer a causal relationship between a quantitative trait (defined as *exposure*) and a disease (*outcome*). The genetic variants associated with a quantitative trait are used as instrumental variables (Ivs) to test the causal relationship between exposure and outcome. Critically, because they are constant, they are not affected by reverse causation and/or confounders. Like the methods previously described, this approach is essentially based on the summary statistics for a set of Ivs chosen to satisfy specific hypotheses, such as the association with exposures, to which appropriate statistical regression methods are applied (39–41). The increasing availability of large datasets and the consequent increasing number of variants that can be tested are facilitating the application of the MR approach.

Another limiting factor in making causal inferences about the involvement of a given immune cell in a particular autoimmune disease is the relatively small sample size of the immune cell GWAS performed to date, which constrains the generation of robust instrumental variables for Mendelian randomization approaches. Furthermore, the true disease-related cell type may not even have been assessed in immune cell trait GWAS! The latter limitations can be overcome thanks to the development of more advanced cytofluorimeters and the implementation of automation methods to permit considerable enlargement of the immune-phenotypic space (**Figure 5D**) examined in an increasingly larger number of individuals.

THERAPEUTIC TARGETS, MULTI-SPECIFICITY, AND PERSONALIZED MEDICINE

After establishing co-localized association signals between immune cell traits and autoimmune disease risk that are likely to share a causal variant pointed by Mendelian randomization approaches, a critical step toward the identification of the right therapeutic targets is to identify the DNA variant, and establish/infer the protein product, underpinning such overlapping associations that could be modulated therapeutically.

In short, an initial strategy commonly applied to statistically exclude all but ideally one or a few polymorphisms as causal variants in GWAS-associated regions encompasses several methods known collectively as “fine mapping” (42). This strategy requires an unbiased, and as comprehensive as possible, ascertainment of genetic variation -through large-scale DNA sequencing and the use of informative imputation panels- to split the genetic contributions of individual variants in an associated region, allowing prioritization of those with the highest probability of being causal. The most plausible causal polymorphisms present in the so-called “credible set” are then ranked using several metrics, including sequence conservation across species and functional genomic data (such as transcription factor binding), which produce a score predicting functional relevance. Unfortunately, even after these methods have been applied, the genetic resolution of association signals to a single-variant, single-gene may still be limited by several factors. These include the strong linkage disequilibrium (non-random association of alleles at different loci in a given population) (43) between several candidate variants that in extreme cases may be so closely related as to be genetically indistinguishable (because they always co-occur in the same individuals). An additional difficulty which hampers variant functional annotation, arises from the fact that the vast majority (~80%) of lead variants of association signals with immune traits are localized in “non-coding regions” of the genome with only a fraction of them altering known sequence motifs of transcription factors (3, 10, 44), thus not easy to interpret, even though they must play a very relevant role in gene expression regulation. Most importantly, even statistical refinement of the association signal to a single putative causative DNA variant does not in itself indicate that the gene harboring is causative. In fact, there are multiple examples of long-range control of gene expression by variants located in neighboring genes detected through technologies such as promoter capture with “Hi-C” (45).

Still, despite these difficulties, the identification of the causal genes highlighting their products as therapeutic targets can be often achieved through expression quantitative trait loci (eQTLs, based on the analysis of the influence of genetic variation on RNA levels) and/or protein QTLs (pQTLs, based on the analysis of the influence of genetic variation on protein levels), which in the cytofluorimetric studies are represented by the expression level of immune cell protein levels (MFIs). In addition to *cis* effects, these analyses can reveal *trans* effects, i.e., trans pQTL and eQTL associations that highlight protein targets for therapeutic intervention encoded by genes located far away, and even on different chromosomes, from the variant/gene underlying the primary association signal but whose expression is affected by it or its protein product or a nearby genetically related variant.

The utility of pQTL and eQTL analyses extends to the determination of the effective direction of the association. This is inferred from the direction of change in levels of gene products associated with disease risk – for example, evaluating whether a disease-protective allele (whose effect we want to therapeutically reproduce) decreases or increases transcript levels of a gene or corresponding protein. This is thus a critical step because it informs the direction (inhibition/stimulation) of therapeutic modulation of the target.

Such analyses are facilitated by the rapidly growing number of large datasets annotating information that can systematically help to bridge GWAS associations to expression levels. One key resource is the Genotype-Tissue Expression (GTEx) catalogue, providing eQTL analysis for 49 human tissues in 838 individuals (46). Additional sources to help assess the impact of regulatory variants include databases, such as the Human Induced Pluripotent Stem Cell Initiative (HipSci) (47) reporting mutations in reprogrammed induced pluripotent stem cell and

LINkage Disequilibrium-based Annotation, LinDA brower (http://linda.irgb.cnr.it) that provides annotations and statistics for the query variant and for variants in linkage disequilibrium with the query. While these comprehensive public resources to study tissue-specific transcription and expression are essential to identify target genes and direction of effects of associations signals with immune cells and other trait types, GWAS results may in turn give rise to more targeted studies of transcription and regulation to elucidate the fine mechanisms of gene expression at specific loci. As an example, in a GWAS analysis it was uncovered the association of multiple sclerosis and systemic lupus erythematosus with a genetic variant in the 3'UTR of the *TNFSF13B* gene, which encodes the cytokine B-cell-activating-factor (BAFF) (10). The same signal also correlated with increased circulating B cell and immunoglobulin levels, giving a potential mechanistic explanation for the disease association. The causal variant underlying these associations was found to be an insertion-deletion (GCTGT > A, [GCTG/-] where the minor risk-associated allele A (referred as 'BAFF-var') was predicted to create an upstream alternative polyadenylation site (APA). This APA was experimentally demonstrated and the resulting shorter transcript, BAFF-var mRNA, was more actively translated than the long wild-type mRNA (BAFF-WT) partly because it lacked a site of repression by microRNA miR-15a (10). Subsequent analyses showed that the short 3'UTR lacked also a binding site of repression by the RNA Binding Protein (RBP) NF90 and revealed that, in the BAFF-WT mRNA, NF90 suppresses BAFF production by promoting the interaction of miR-15a with BAFF-WT mRNA. As a consequence of this lack of repression of BAFF expression due to BAFF-var, soluble BAFF is produced at higher levels determining a cascade of immune events leading to increased risk for systemic lupus erythematosus and multiple sclerosis (48). It is expected that this type of fine analysis of the regulation of gene expression will increasingly contribute to a detailed understanding of the molecular mechanisms of genetic associations with immune traits.

The obvious next critical step toward the therapeutic modulation of a protein target identified with genetic approaches is the assessment of its druggability – that is, its susceptibility to be potentially modulated in its effects by drug-like small molecules (typically targeting hydrophobic pockets) or by so called "biologicals" (more commonly targeting extracellular domains such as those of receptor proteins or soluble molecules) or by new molecular approaches, such as those based on small interfering RNA, antisense oligonucleotides, mRNA delivery, gene editing with CRISPR–Cas9, and PROteolysis-TArgeting Chimaeras (PROTAC) (49–51).

In particular, the protein target identification approach presented here, built on the results of flow cytometry coupled with genetic data, offers an obvious opportunity for therapeutic intervention through the generation of biological products, specifically, as we detailed below, through a new class of poly-specific antibodies. In contrast, many current monospecific antibody-based therapies aimed to block, or in few cases enhance, the activity of a single antigen generally expressed on the cell surface membrane, such as anti-CD28, CD40, and CD25. Nevertheless, these mono-specific drugs are affected by poor cell specificity causing reduced efficacy and predisposition to side effects like increased risk of other autoimmune diseases. Indeed, targeting broadly expressed markers such as CD25 or CD27, which are expressed in both T and B cells, or CD28, expressed in both CD4 and CD8 T cells, could cause unspecific blocking of this marker in cells that are not involved in a specific disease (3).

For instance, IL2RA, also known as CD25, encodes the alpha chain of IL-2 receptor and is expressed in regulatory T cells (Tregs), activated effector T cells, but also in B cells. In 2013, measuring CD25 in T cells only and using about 8.2 million variants, an overlapping association between CD25hi effector T cells and type 1 diabetes was found in the *IL2RA* region (1). More recently, by increasing both the cell types where CD25 has been assessed and the number of interrogated variants, seven independent signals in the *IL2RA* locus (all regulating CD25 expression) were identified (3) (**Figure 6**). Some signals were T cell specific; others were B cell specific, still, others involved both T and B cells. Four out of the seven independent signals detected in this region were associated with immune-diseases and pointed to different traits, in some cases with opposite direction of effect, potentially leading to adverse therapeutic complications (**Table 1**). In more details, the inhibition of T cells expressing high levels of CD25 may be efficacious in Crohn's disease, but harmful in type 1 diabetes and juvenile rheumatoid arthritis for which a stimulation of the same cells is likely to be effective. These data also suggest that reduction of CD25 on naive effector helper T cells could be an effective therapy in multiple sclerosis and alopecia areata. But inhibition of CD25 on a specific subset of memory B cells called late memory B cells (identified as positive for CD19, but negative for IgD and CD38) could be useful in vitiligo and autoimmune thyroiditis therapies (**Figure 6**).

Overall, the genetic associations observed in the *IL2RA* region can predict the efficacy and potential adverse effect of the broad blocking of CD25 that causes a reduction of CD25 activity in cells not implicated in disease predisposition (e.g., Tregs).

A similar scenario was observed for several antigens, such as CD32, CD28, and CD40, whose increase is associated with predisposition to some diseases, but also with protection from others (**Table 1**). But if the targeted antigen can be addressed in specific cells (for instance either B or T cells in the case of CD25), the adverse effects should be minimized. Thus, the generation of multi-specific drugs, able to recognize more than one antigen simultaneously, can provide an optimal way to ensure specificity and reduce adverse effects.

Multi-specific drugs are in clinical trials especially for cancer treatment, where an antibody binds immune cells such as CD3-positive, while another antibody binds cancer cells, thereby redirecting T-cell cytotoxicity to malignant cells (52, 53). However, the same approach can be useful to engage two molecules on the membrane of one cell (*in-cis* binding). For instance, MGD010 is a dual-affinity retargeting (DART) protein which simultaneously binds the B cell surface proteins CD32B and CD79B to deliver a co-inhibitory signal that dampens B cell activation (54). The intended mechanism of MGD010 is to modulate the function of human B cells while avoiding their depletion and could be useful for treatment of rheumatoid arthritis

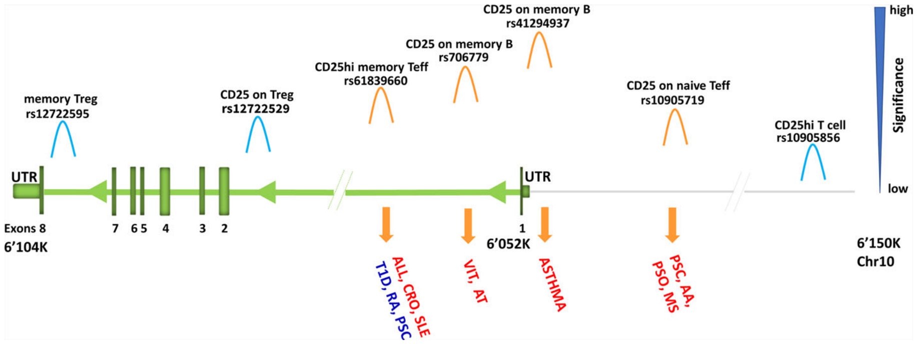


FIGURE 6 | Association signals at *IL2RA* region. Representation of *IL2RA* gene (green) and about 100 kb upstream to the gene (grey line). The association signals with immune cell traits are depicted by «hills» which are colored in orange or light blue if overlapping or not overlapping with disease-association signals, respectively. Disease is in red if the predisposing allele is associated with increase of immune cell traits, whereas it is in blue if the predisposing allele is associated with decrease of immune cell traits. Disease acronyms: T1D, type one diabetes; RA, rheumatoid arthritis; PSC, primary sclerosis cholangitis; ALL, allergy; CRO, Crohn's disease; SLE, systemic lupus erythematosus; VIT, vitiligo; AT, autoimmune thyroiditis; AA, alopecia areata; PSO, psoriasis; MS, multiple sclerosis.

and other autoimmune and inflammatory diseases. Notwithstanding, only few bispecific antibodies have been approved and marketed, namely blinatumomab (55), simultaneously targeting B cell CD19 antigen and T cell CD3 antigen against B cell malignancies, and emicizumab (56, 57), targeting coagulation factors IXa and X against hemophilia A. Finally, catumaxomab, approved in Europe for the intraperitoneal treatment of malignant ascites, binds to the epithelial cell adhesion molecule (EpCAM), T cells (*via* CD3), and to accessory cells, including dendritic cells, macrophages, and natural killer cells through its Fc-fragment (58). Approved in 2009, catumaxomab was however withdrawn from the US market in 2013 and from European market in 2017 when the company became insolvent.

More recently, tri-specific drugs have been developed. Among them, a single molecule designed by Xu and colleagues is able to bind three HIV-1 envelope determinants: the CD4 binding site, the membrane proximal external region, and the V1V2 glycan site, showing higher potency and breadth compared to previously used antibodies and complete immunity against a mixture of simian-human immunodeficiency viruses (SHIVs) in nonhuman primates (59).

The reason why few multi-specific drugs have been created and approved is that they are not easy to generate due to their instability, low solubility, unwanted inter-subunit associations, and enhanced immunogenicity (60). The evaluation of these therapeutic properties as well as manufacturability and safety profile is called developability.

Another relevant consideration is the choice of the most appropriate dose. Several studies demonstrated that even when a drug is able to ameliorate a disease condition, its administration at a wrong concentration can cause potentially deadly side-effects. This happened in 2006 when six healthy young males were enrolled in the first phase 1 clinical trial of the CD28 super-agonist TGN1412, which can activate T cells, particularly regulatory T cells, thus potentially efficacious against autoimmunity where a reduced function of Tregs is expected. All volunteers had an unpredicted multiple cytokines release syndrome and underwent intensive cardiopulmonary support, dialysis, and administration of both a high-dose of anti-inflammatory drugs such as methylprednisolone and an anti–interleukin-2 receptor antagonist antibody. Fortunately, all six volunteers survived (61). It was clear that the drug activated effector T cells instead of Tregs. Some years later, the reasons for the preclinical study failure of TGN1412 were found (62). Firstly, only about 2% of T cells circulate in the peripheral blood (63), thus human T cells used for *in vitro* studies (which derive from that 2%) respond differently compared to those *in vivo*, which include also the remaining 98% of T cells. Secondly, in mouse models living in a germ-free animal house used to test the drug, CD4 effector memory cells are much lower in numbers and easily controllable by TGN1412-activated Tregs compared to humans. Thirdly, in cynomolgus macaques, also used to test the drug, CD4 effector memory cells down-regulate CD28, and thus it cannot bind TGN1412; this does not occur in humans. In 2014, Tabares and colleagues (64) demonstrated that a strong reduction of TGN1412, now renamed TAB08, accompanied by the administration of corticosteroid drug (such as methylprednisolone), activates Tregs without a cytokine storm, thus making it useful in rheumatoid arthritis and other autoimmune therapies.

The TGN1412 results exemplify the need to identify the correct dose for the correct target. Notably, the proper dose could also depend on our genome, indeed, differences in our DNA sequence that affect the levels of the drug target (such as specific cell type or protein) could modify the efficacy of the pharmacological treatments, thus an individual could need a different drug concentration compared to another individual - a type of personalized medicine.

CONCLUDING REMARKS

Flow cytometry combined with systematic GWAS of immune traits in general population cohorts and case-control GWAS data on autoimmune disease risk is a powerful strategy to identify specific proteins, cells, and pathways involved in the

TABLE 1 | Immune traits associated with diseases *via* overlapping genetic association and having opposite direction of effect in different diseases. Extracted from Orrù et al., 2020 (3).

Cell trait name	Primary drug targets	Disease	Proposed therapeutic modulation of primary drug targets	Expected increased risk for other autoimmune disease (side effect)
CD32 on monocyte	CD32	CRO, IBD, KD, AS, UC	inhibition	SLE
CD32 on monocyte	CD32	SLE	activation	CRO, IBD, KD, AS, UC
CD28 on CD39+ CD4+	CD28	UC, CEL	activation	MS
CD28 on CD39+ CD4+	CD28	MS	inhibition	UC, CEL
CD28 on CD4+	CD28	UC, CEL	activation	MS
CD28 on CD4+	CD28	MS	inhibition	CEL, UC
CCR2 on monocyte	CCR2	BD, CEL	activation	SLE
CCR2 on monocyte	CCR2	SLE	inhibition	CEL, BD
HLA DR on CD14- CD16+ monocyte	HLA DR	CEL, Allergy, MS, Cutaneous squamous cell carcinoma	inhibition	VIT, AA
HLA DR on CD14- CD16+ monocyte	HLA DR	VIT, AA	activation	CEL, Allergy, MS, Cutaneous squamous cell carcinoma
CD80 on myeloid DC (especially CD62L+)	CD80	CEL	inhibition	CRO, IBD, Allergy
CD80 on myeloid DC (especially CD62L+)	CD80	CRO, IBD, Allergy	activation	CEL
CD45RA on naive CD4+	CD45RA	Allergy, MS	inhibition	RA
CD45RA on naive CD4+	CD45RA	RA	activation	MS, Allergy
CD25hi%CD4+ (especially CD25hi CD45RA- CD4 not Treg %CD4+)	CD25, CD4, CD3	T1D, PSC, JIA	activation	Allergy, CRO
CD25hi%CD4+ (especially CD25hi CD45RA- CD4 not Treg %CD4+)	CD25, CD4, CD3	Allergy, CRO	inhibition	T1D, PSC, JIA
CD25 on CD45RA- CD4 not Treg	CD25	T1D, PSC, JIA	activation	Allergy, CRO
CD25 on CD45RA- CD4 not Treg	CD25	Allergy, CRO	inhibition	T1D, PSC, JIA
CD25 on CD4+	CD25	T1D, PSC, JIA	activation	Allergy, CRO
CD25 on CD4+	CD25	Allergy,CRO	inhibition	T1D, PSC, JIA
CD25++ CD8br%Tcells	CD25, CD8	T1D, PSC, JIA	activation	Allergy, CRO
CD25++ CD8br%Tcells	CD25, CD8	Allergy, CRO	inhibition	T1D, PSC, JIA
CD11c on myeloid DC	CD11c	IgAN	activation	SLE
CD11c on myeloid DC	CD11c	SLE	inhibition	IgAN
CD19 on B cell (especially sw mem IgD-CD27+)	CD19	IBD, CRO, Ob	activation	T1D
CD19 on B cell (especially sw mem IgD-CD27+)	CD19	T1D	inhibition	CRO, IBD, Ob
IgD+ AC	IgD, CD19/CD20	Allergy, Asthma, ALL, AM	inhibition	CC, CRO, PBC, RA, T1D, UC, Bronchial hyperresponsiveness in asthma, Selective IgA deficiency, Liver biliary cirrhosis
IgD+ AC	IgD, CD19/CD20	CC, CRO, PBC, RA, T1D, UC, Bronchial hyperresponsiveness in asthma, Selective IgA deficiency, Liver biliary cirrhosis	activation	Allergy, Asthma, ALL, AM
Unsw Mem (IgD+CD27+) %lymphocyte	IgD, CD27, CD19/CD20	HBVI, CRO, IBD, MS, SLE	inhibition	RA, Depression, KD
Unsw Mem (IgD+CD27+) %lymphocyte	IgD, CD27, CD19/CD20	RA, Depression, KD	activation	HBVI, CRO, IBD, MS, SLE
CD27 on memory B cell (especially IgD-CD38dim)	CD27	HBVI, CRO, IBD, MS, SLE	inhibition	RA, KD
CD27 on memory B cell (especially IgD-CD38dim)	CD27	RA, KD	activation	HBVI, CRO, IBD, MS, SLE
CD40 on B cell (especially IgD-CD27-)	CD40	HBVI, CRO, IBD, MS, SLE	activation	RA, KD
CD40 on B cell (especially IgD-CD27-)	CD40	RA, KD	inhibition	HBVI, CRO, IBD, MS, SLE
IgD- CD27- %B cell	CD19/CD20	HBVI, CRO, IBD, MS, SLE	activation	RA, KD
IgD- CD27- %B cell	CD19/CD20	RA, KD	inhibition	HBVI, CRO, IBD, MS, SLE

etiopathogenesis of immune related diseases. After appropriate validation with functional studies, this strategy will be increasingly relevant to identify therapeutic targets and reinforce causal relationships as the technology will evolve to permit a considerable expansion of the number of markers assessed simultaneously by flow cytometry and of the sample size of the studies. Corresponding advances in the generation of a new class of *in-cis*, multi-specific antibodies to engage these targets will progressively increase efficacy and minimize the potential side effects in the treatment of autoimmune diseases.

In summary, from flow cytometry data collection to drug therapy development, four main steps are relevant:

- coupling flow cytometry data to genetics in the general population sample set to identify the genetic component driving the interindividual immune variability;
- systematically searching for overlapping association between immune trait-associated variants (from population-based datasets) and disease-associate variants (from case-control datasets);
- causality confirmation of identified disease risk variants through functional studies;
- drug development in a cell-specific context.

AUTHOR CONTRIBUTIONS

All authors contributed to the article and approved the submitted version.

FUNDING

The preparation of this manuscript was supported by the Intramural Research Program of the National Institute on Aging, National Institutes of Health (NIH) (contract HHSN271201600005C), the Horizon 2020 Research and Innovation Program of the European Union (grant 633964), and Fondazione di Sardegna (grant U1301.2015/AI.1157. BE Prat. 2015-1651).

ACKNOWLEDGMENTS

We thank Marcella Devoto, David Schlessinger, and Gianluca Rotta for their precious advice during the drafting of the manuscript.

REFERENCES

1. Orrù V, Steri M, Sole G, Sidore C, Virdis F, Dei M, et al. Genetic Variants Regulating Immune Cell Levels in Health and Disease. *Cell* (2013) 155 (1):242–56. doi: 10.1016/j.cell.2013.08.041
2. Mangino M, Roederer M, Beddall MH, Nestle FO, Spector TD. Innate and Adaptive Immune Traits Are Differentially Affected by Genetic and Environmental Factors. *Nat Commun* (2017) 8:13850. doi: 10.1038/ncomms13850
3. Orrù V, Steri M, Sidore C, Marongiu M, Serra V, Olla S, et al. Complex Genetic Signatures in Immune Cells Underlie Autoimmunity and Inform Therapy. *Nat Genet* (2020) 52(11):1266. doi: 10.1038/s41588-020-00718-6
4. Roederer M, Quaye L, Mangino M, Beddall MH, Mahnke Y, Chattopadhyay P, et al. The Genetic Architecture of the Human Immune System: A Bioresource for Autoimmunity and Disease Pathogenesis. *Cell* (2015) 161 (2):387–403. doi: 10.1016/j.cell.2015.02.046
5. Patin E, Hasan M, Bergstedt J, Rouilly V, Libri V, Urrutia A, et al. Natural Variation in the Parameters of Innate Immune Cells Is Preferentially Driven by Genetic Factors. *Nat Immunol* (2018) 19(3):302–14. doi: 10.1038/s41590-018-0049-7
6. Akbari P, Vuckovic D, Jiang T, Kundu K, Kreuzhuber R, Bao EL, et al. Genetic Analyses of Blood Cell Structure for Biological and Pharmacological Inference. *bioRxiv preprint* (2020). doi: 10.1101/2020.01.30.927483
7. Okada D, Nakamura N, Setoh K, Kawaguchi T, Higasa K, Tabara Y, et al. Genome-Wide Association Study of Individual Differences of Human Lymphocyte Profiles Using Large-Scale Cytometry Data. *J Hum Genet* (2021) 66(6):557–67. doi: 10.1038/s10038-020-00874-x
8. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* (2017) 169(7):1177–86. doi: 10.1016/j.cell.2017.05.038
9. Cook D, Brown D, Alexander R, March R, Morgan P, Satterthwaite G, et al. Lessons Learned From the Fate of Astrazeneca's Drug Pipeline: A Five-Dimensional Framework. *Nat Rev Drug Discov* (2014) 13(6):419–31. doi: 10.1038/nrd4309
10. Steri M, Orrù V, Idda ML, Pitzalis M, Pala M, Zara I, et al. Overexpression of the Cytokine BAFF and Autoimmunity Risk. *N Engl J Med* (2017) 376 (17):1615–26. doi: 10.1056/NEJMoa1610528
11. Picot J, Guerin CL, Le Van Kim C, Boulanger CM. Flow Cytometry: Retrospective, Fundamentals and Recent Instrumentation. *Cytotechnology* (2012) 64:109–30. doi: 10.1007/s10616-011-9415-0
12. Marti GE, Rawstron AC, Ghia P, Hillmen P, Houlston RS, Kay N, et al. Diagnostic Criteria for Monoclonal B-Cell Lymphocytosis. *Br J Haematol* (2005) 130:325–32. doi: 10.1111/j.1365-2141.2005.05550.x
13. Pina-Vaz C, Costa-de-Oliveira S, Silva-Dias A, Silva AP, Teixeira-Santos R, Rodrigues AG. "Flow Cytometry in Microbiology: The Reason and the Need". In: J Robinson, A Cossarizza, editors. *Single Cell Analysis. Series in Bioengineering*. Singapore: Springer (2017). doi: 10.1007/978-981-10-4499-1_7
14. Wilkinson MG. *Flow Cytometry in Microbiology*. Ireland: Wilkinson Department of Life Sciences, University of Limerick (2015). p. 230.
15. Goddard GR, Sanders CK, Martin JC, Kaduchak G, Graves SW. Analytical Performance of an Ultrasonic Particle Focusing Flow Cytometer. *Anal Chem* (2007) 79(22):8740–6. doi: 10.1021/ac071402t
16. Bandura DR, Baranov VI, Ornatsky OI, Antonov A, Kinach R, Lou X, et al. Mass Cytometry: Technique for Real Time Single Cell Multitarget Immunoassay Based on Inductively Coupled Plasma Time-of-Flight Mass Spectrometry. *Anal Chem* (2009) 81(16):6813–22. doi: 10.1021/ac901049w
17. Payne K, Li W, Salomon R, Ma CS. OMIP-063: 28-Color Flow Cytometry Panel for Broad Human Immunophenotyping. *Cytometry A* (2020) 97 (8):777–81. doi: 10.1002/cyto.a.24018
18. Qiu P, Simonds EF, Bendall SC, Gibbs KDJr., Bruggner RV, Linderman MD, et al. Extracting a Cellular Hierarchy From High-Dimensional Cytometry Data With SPADE. *Nat Biotechnol* (2011) 29(10):886–91. doi: 10.1038/nbt.1991
19. Van der Maaten L, Hinton G. Visualizing Data Using T-Sne. *J Mach Learn Res* (2008) 9:2579–605.
20. Shekhar K, Brodin P, Davis MM, Chakraborty AK. Automatic Classification of Cellular Expression by Nonlinear Stochastic Embedding (ACCENSE). *Proc Natl Acad Sci USA* (2014) 111:202–7. doi: 10.1073/pnas.1321405111
21. Becher B, Schlitzer A, Chen J, Mair F, Sumatoh HR, Teng KWW, et al. High-Dimensional Analysis of the Murine Myeloid Cell System. *Nat Immunol* (2014) . 15:1181–9. doi: 10.1038/ni.3006
22. Amir E-AD, Davis KL, Tadmor MD, Simonds EF, Levine JH, Bendall SC, et al. Visne Enables Visualization of High Dimensional Single-Cell Data and Reveals Phenotypic Heterogeneity of Leukemia. *Nat Biotechnol* (2013) 31:545–52. doi: 10.1038/nbt.2594
23. Levine JH, Simonds EF, Bendall SC, Davis KL, Amir E-AD, Tadmor MD, et al. Data-Driven Phenotypic Dissection of AML Reveals Progenitor-Like Cells That Correlate With Prognosis. *Cell* (2015) 162:184–97. doi: 10.1016/j.cell.2015.05.047

24. Bendall SC, Davis KL, Amir E-AD, Tadmor MD, Simonds EF, Chen TJ. Single-Cell Trajectory Detection Uncovers Progression and Regulatory Coordination in Human B Cell Development. *Cell* (2014) 157:714–25. doi: 10.1016/j.cell.2014.04.005
25. Mair F, Hartmann FJ, Mrdjen D, Tosevski V, Krieg C, Becher B. The End of Gating? An Introduction to Automated Analysis of High Dimensional Cytometry Data. *Eur J Immunol* (2016) 46:34–43. doi: 10.1002/eji.201545774
26. Bruggner RV, Bodenmiller B, Dill DL, Tibshirani RJ, Nolan GP. Automated Identification of Stratifying Signatures in Cellular Subpopulations. *Proc Natl Acad Sci USA* (2014) 111:E2770–7. doi: 10.1073/pnas.1408792111
27. Schuh E, Berer K, Mulazzani M. Features of Human CD3+CD20+ T Cells. *J Immunol* (2016) 197(4):1111–7. doi: 10.4049/jimmunol.1600089
28. Roederer M. Compensation in Flow Cytometry. *Curr Protoc Cytom* (2002) 22:1.14.1–1.14.20. doi: 10.1002/0471142956.cy0114s22
29. Fung E, Esposito L, Todd JA, Wicker LS. Multiplexed Immunophenotyping of Human Antigen-Presenting Cells in Whole Blood by Polychromatic Flow Cytometry. *Nat Protoc* (2010) 5(2):357–70. doi: 10.1038/nprot.2009.246
30. van Velzen JF, Laros-van Gorkom BA, Pop GA, van Heerde WL. Multicolor Flow Cytometry for Evaluation of Platelet Surface Antigens and Activation Markers. *Thromb Res* (2012) 130(1):92–8. doi: 10.1016/j.thromres.2012.02.041
31. Berny-Lang MA, Frelinger ALIII, Barnard MR, Michelson AD. "Flow Cytometry". In: *Chaper 29 Platelets*, *3rd edition*. Elsevier (2013).
32. Chan RC, Kotner JS, Chuang CM, Gaur A. Stabilization of Pre-Optimized Multicolor Antibody Cocktails for Flow Cytometry Applications. *Cytometry Part B (Clinical Cytometry)* (2017) 92(6):508–24. doi: 10.1002/cyto.b.21371
33. Kadić E, Moniz RJ, Huo Y, Chi A, Kariv I. Effect of Cryopreservation on Delineation of Immune Cell Subpopulations in Tumor Specimens as Determinated by Multiparametric Single Cell Mass Cytometry Analysis. *BMC Immunol* (2017) 18:6. doi: 10.1186/s12865-017-0192-1
34. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian Test for Colocalisation Between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet* (2014) 10(5):e1004383. doi: 10.1371/journal.pgen.1004383
35. Paaby AB, Rockman MV. The Many Faces of Pleiotropy. *Trends Genet* (2013) 29(2):66–73. doi: 10.1016/j.tig.2012.10.010
36. Ferreira MA, Mangino M, Brumme CJ, Zhao ZZ, Medland SE, Wright MJ, et al. Quantitative Trait Loci for CD4:CD8 Lymphocyte Ratio Are Associated With Risk of Type 1 Diabetes and HIV-1 Immune Control. *Am J Hum Genet* (2010) 86(1):88–92. doi: 10.1016/j.ajhg.2009.12.008
37. Aguirre-Gamboa R, Joosten I, Urbano PCM, van der Molen RG, van Rijssen E, van Cranenbroek B, et al. Differential Effects of Environmental and Genetic Factors on T and B Cell Immune Traits. *Cell Rep* (2016) 17(9):2474–87. doi: 10.1016/j.celrep.2016.10.053
38. Lagou V, Garcia-Perez JE, Smets I, Van Horebeek L, Vandebergh M, Chen L, et al. Genetic Architecture of Adaptive Immune System Identifies Key Immune Regulators. *Cell Rep* (2018) 25(3):798–810.e6. doi: 10.1016/j.celrep.2018.09.048
39. Paternoster L, Tilling K, Smith GD. Genetic Epidemiology and Mendelian Randomization for Informing Disease Therapeutics: Conceptual and Methodological Challenges. *PLoS Genet* (2017) 13(10):e1006944. doi: 10.1371/journal.pgen.1006944
40. Davey Smith G, Hemani G. Mendelian Randomization: Genetic Anchors for Causal Inference in Epidemiological Studies. *Hum Mol Genet* (2014) 23(R1): R89–98. doi: 10.1093/hmg/ddu328
41. van der Graaf A, Claringbould A, Rimbert ABIOS Consortium, , Westra H, Li Y, et al. Mendelian Randomization While Jointly Modeling Cis Genetics Identifies Causal Relationships Between Gene Expression and Lipids. *Nat Commun* (2020) 11(1):4930. doi: 10.1038/s41467-020-18716-x
42. Broekema RV, Bakker OB, Jonkers IH. A Practical View of Fine-Mapping and Gene Prioritization in the Post-Genome-Wide Association Era. *Open Biol* (2020) 10(1):190221. doi: 10.1098/rsob.190221
43. Montgomery S. Linkage Disequilibrium — Understanding the Evolutionary Past and Mapping the Medical Future. *Nat Rev Genet* (2008) 9(6):477–85. doi: 10.1038/nrg2361
44. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and Epigenetic Fine Mapping of Causal Autoimmune Disease Variants. *Nature* (2015) 518(7539):337–43. doi: 10.1038/nature13835
45. Schoenfelder S, Javierre BM, Furlan-Magaril M, Wingett SW, Fraser P. Promoter Capture Hi-C: High-Resolution, Genome-Wide Profiling of Promoter Interactions. *J Vis Exp* (2018) 136):57320. doi: 10.3791/57320
46. The GTEx Consortium. The Gtex Consortium Atlas of Genetic Regulatory Effects Across Human Tissues. *Science* (2020) 369(6509):1318–30. doi: 10.1126/science.aaz1776
47. Streeter I, Harrison PW, Faulconbridge AThe HipSci Consortium, , Flicek P, Parkinson H, et al. The Human-Induced Pluripotent Stem Cell Initiative —-Data Resources for Cellular Genetics. *Nucleic Acids Res* (2017) 45(D1): D691–7. doi: 10.1093/nar/gkw928
48. Idda ML, Lodde V, McClusky WG, Martindale JL, Yang X, Munk R, et al. Cooperative Translational Control of Polymorphic BAFF by NF90 and Mir-15a. *Nucleic Acids Res* (2018) 46(22):12040–51. doi: 10.1093/nar/gky866
49. Liang F, Lindgren G, Lin A, Thompson EA, Ols S, Röhss J, et al. Efficient Targeting and Activation of Antigen-Presenting Cells in Vivo After Modified Mrna Vaccine Administration in Rhesus Macaques. *Mol Ther* (2017) 25 (12):2635–47. doi: 10.1016/j.ymthe.2017.08.006
50. Bosley KS. *Nat Rev Drug Discov* (2017) 16:672–3. doi: 10.1038/nrd.2017.191
51. Lai AC, Crews CM. Induced Protein Degradation: An Emerging Drug Discovery Paradigm. *Nat Rev Drug Discov* (2017) 16(2):101–14. doi: 10.1038/nrd.2016.211
52. Dahlén E, Veitonmäki N, Norlén P. Bispecific Antibodies in Cancer Immunotherapy. *Ther Adv Vaccines Immunother* (2018) 6(1):3–17. doi: 10.1177/2515135518763280
53. Brinkmann U, Kontermann RE. The Making of Bispecific Antibodies. *MABS* (2017) 9(2):182–212. doi: 10.1080/19420862.2016.1268307
54. Veri MC, Burke S, Huang L, Li H, Gorlatov S, Tuaillon N, et al. Therapeutic Control of B Cell Activation *via* Recruitment of Fcgamma Receptor Iib (CD32B) Inhibitory Function With a Novel Bispecific Antibody Scaffold. *Arthritis Rheumatol* (2010) 62(7):1933–43. doi: 10.1002/art.27477
55. Bargou R, Leo E, Zugmaier G, Klinger M, Goebeler M, Knop S, et al. Tumor Regression in Cancer Patients by Very Low Doses of a T Cell-Engaging Antibody. *Science* (2008) 321(5891):974–7. doi: 10.1126/science.1158545
56. Kitazawa T, Igawa T, Sampei Z, Muto A, Kojima T, Soeda T, et al. A Bispecific Antibody to Factors Ixa and X Restores Factor VIII Hemostatic Activity in a Hemophilia a Model. *Nat Med* (2012) 18(10):1570–4. doi: 10.1038/nm.2942
57. Knight T, Callaghan MU. The Role of Emicizumab, a Bispecific Factor Ixa- and Factor X-Directed Antibody, for the Prevention of Bleeding Episodes in Patients With Hemophilia a. *Ther Adv Hematol* (2018) 9(10):319–34. doi: 10.1177/2040620718799997
58. Sebastian M. Review of Catumaxomab in the Treatment of Malignant Ascites. *Cancer Manag Res* (2010) 2:283–6. doi: 10.2147/CMR.S14115
59. Xu L, Pegu A, Rao E, Doria-Rose N, Beninga J, McKee K, et al. Trispecific Broadly Neutralizing HIV Antibodies Mediate Potent SHIV Protection in Macaques. *Science* (2017) 358(6359):85–90. doi: 10.1126/science.aan8630
60. Sawant MS, Streu CN, Wu L, Tessier PM. Toward Drug-Like Multispecific Antibodies by Design. *Int J Mol Sci* (2020) 21(20):7496. doi: 10.3390/ijms21207496
61. Suntharalingam G, Perry MR, Ward S, Brett SJ, Castello-Cortes A, Brunner MD, et al. Cytokine Storm in a Phase 1 Trial of the Anti-CD28 Monoclonal Antibody Tgn1412. *N Engl J Med* (2006) 355(10):1018–28. doi: 10.1056/NEJMoa063842
62. Hunig T. The Rise and Fall of the CD28 Superagonist TGN1412 and Its Return as TAB08: A Personal Account. *FEBS J* (2016) 283(18):3325–34. doi: 10.1111/febs.13754
63. Ganusov VV, De Boer RJ. Do Most Lymphocytes in Humans Really Reside in the Gut? *Trends Immunol* (2007) 28(12):514–8. doi: 10.1016/j.it.2007.08.009
64. Tabares P, Berr S, Romer PS, Chuvpilo S, Matskevich AA, Tyrsin D, et al. Human Regulatory T Cells Are Selectively Activated by Low-Dose Application of the CD28 Superagonist TGN1412/TAB 08. *Eur J Immunol* (2014) 44(4):1225–36. doi: 10.1002/eji.201343967

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Immunology

ORIGINAL RESEARCH
published: 08 September 2021
doi: 10.3389/fimmu.2021.665160

Shared Genetic Liability Between Major Depressive Disorder and Atopic Diseases

Hongbao Cao[1,2†], *Sheng Li*[3,4†], *Ancha Baranova*[2,5†] and *Fuquan Zhang*[6,7*]

[1] Department of Psychiatry, First Hospital/First Clinical Medical College of Shanxi Medical University, Taiyuan, China, [2] School of Systems Biology, George Mason University, Fairfax, VA, United States, [3] Institute of Systems Medicine, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China, [4] Suzhou Institute of Systems Medicine, Suzhou, China, [5] Research Centre for Medical Genetics, Moscow, Russia, [6] Department of Psychiatry, The Affiliated Brain Hospital of Nanjing Medical University, Nanjing, China, [7] Institute of Neuropsychiatry, The Affiliated Brain Hospital of Nanjing Medical University, Nanjing, China

OPEN ACCESS

Edited by:
Lam Cheung Tsoi,
University of Michigan, United States

Reviewed by:
Cunyou Zhao,
Southern Medical University, China
Mona-Rita Yacoub,
San Raffaele Hospital (IRCCS), Italy

***Correspondence:**
Fuquan Zhang
zhangfq@njmu.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Autoimmune and Autoinflammatory Disorders, a section of the journal Frontiers in Immunology

Received: *07 February 2021*
Accepted: *16 August 2021*
Published: *08 September 2021*

Citation:
Cao H, Li S, Baranova A, and Zhang F (2021) Shared Genetic Liability Between Major Depressive Disorder and Atopic Diseases. Front. Immunol. 12:665160. doi: 10.3389/fimmu.2021.665160

Objectives: Deciphering the genetic relationships between major depressive disorder (MDD) and atopic diseases (asthma, hay fever, and eczema) may facilitate understanding of their biological mechanisms as well as the development of novel treatment regimens. Here we tested the genetic correlation between MDD and atopic diseases by linkage disequilibrium score regression.

Methods: A polygenic overlap analysis was performed to estimate shared genetic variations between the two diseases. Causal relationships between MDD and atopic diseases were investigated using two-sample bidirectional Mendelian randomization analysis. Genomic loci shared between MDD and atopic diseases were identified using cross-trait meta-analysis. Putative functional genes were evaluated by fine-mapping of transcriptome-wide associations.

Results: The polygenic analysis revealed approximately 15.8 thousand variants causally influencing MDD and 0.9 thousand variants influencing atopic diseases. Among these variants, approximately 0.8 thousand were shared between the two diseases. Mendelian randomization analysis indicates that genetic liability to MDD has a causal effect on atopic diseases ($b = 0.22$, $p = 1.76 \times 10^{-6}$), while genetic liability to atopic diseases confers a weak causal effect on MDD ($b = 0.05$, $p = 7.57 \times 10^{-3}$). Cross-trait meta-analyses of MDD and atopic diseases identified 18 shared genomic loci. Both fine-mapping of transcriptome-wide associations and analysis of existing literature suggest the estrogen receptor β-encoding gene *ESR2* as one of the potential risk factors for both MDD and atopic diseases.

Conclusion: Our findings reveal shared genetic liability and causal links between MDD and atopic diseases, which shed light on the phenotypic relationship between MDD and atopic diseases.

Keywords: major depressive disorder, Mendelian randomization, meta-analyses, asthma, atopic diseases

INTRODUCTION

Mental disorders confer a heavy burden on society (1). Major depressive disorder (MDD), the most prevalent mental disorder accompanied by considerable morbidity, mortality, and risk of suicide, is characterized by persistent low mood (2). MDD and depressive symptoms have close associations with certain physical conditions. Generally speaking, long-term depression adds to the risk for somatic illness, and, vice versa, chronic somatic diseases are frequently accompanied by depression (3). When comorbid with other ailments, for example, atopic diseases (ADs), MDD produces worse clinical outcomes and incurs higher healthcare costs.

ADs are driven by the dysfunction of the immune system. Three kinds of common ADs, namely, asthma, hay fever (allergic rhinitis), and eczema (atopic dermatitis), may coexist in the same individuals (4). Asthma, a chronic airway disease that is common worldwide, is characterized by coughing, wheezing, shortness of breath, and/or chest tightness due to increased airway reactivity, inflammation, and/or mucus production. In 2015, asthma affected 358 million people globally and caused about 400,000 deaths (5). Allergic rhinitis is an inflammatory disease characterized by nasal congestion, rhinorrhea, sneezing, and/or nasal itching. Allergic rhinitis is one of the most common diseases in adults (20%~30%), and the most common chronic disease in children (up to 40%) in the United States (6). Eczema is an inflammatory skin disease that is caused by a dysfunction of a skin barrier followed by aberrant inflammation/immune responses; this disease is affecting 5% of the population worldwide (7). Together, symptoms of ADs significantly impair quality of life and impose a heavy cost on society. Common comorbidities of MDD with ADs have been documented previously (8–12). Specifically, allergic rhinitis has been shown to have a positive association with MDD (odds ratio: 1.24) (8). In patients with asthma, the hazard ratio of MDD increases by 35%, and MDD patients show about 25% increased hazard ratio for being affected by asthma (9). Atopic eczema is also associated with an increased incidence of new depression (hazard ratio: 1.14) (10).

Although previous studies have detected associations between MDD and ADs, several key questions remain pending: 1) to what extent may the two conditions share genetic components? 2) Are the phenotypic associations mediated by genetic variations? 3) What molecular and cellular mechanisms underline these associations?

Genetic relationships between two traits are commonly quantified by genetic correlation coefficients. The sign of the correlation coefficient indicates directions of the shared genetic effects. When dealing with mixtures of effect directions across shared genetic variants, genetic correlation analyses may be underpowered (13). A polygenic overlap was recently proposed to measure the fraction of genetic variants causally associated with both traits over the total number of causal variants across a pair of traits involved (13).

Mendelian randomization (MR) is an analytic framework that utilizes genetic variants as instrumental variables to test for causative association between an exposure and an outcome (14). Recently, a general type of SMR (GSMR) had been developed by leveraging power from multiple genetic variants to account for linkage disequilibrium (LD) between the variants (15).

Recently, Zhu et al. reported a causal effect of MDD on asthma and identified 10 loci shared by asthma and MDD by cross-trait meta-analysis (16). The GWAS dataset for MDD, however, did not include the 23andMe samples. We set on taking this line of investigation further, by both utilizing a larger MDD dataset and including two other ADs related to asthma, namely, allergic rhinitis and atopic dermatitis. Asthma, allergic rhinitis, and atopic dermatitis genetically correlate with each other and are often comorbid (17). The genetic liability to MDD may confer a causal effect on all of these ADs. Dissection of this shared genetic liability may deliver novel insights into the pathophysiology of both MDD and ADs.

METHODS

GWAS Summary Datasets and Quality Control

This study relied on both de-identified publicly available summary-level GWAS data and the pre-approval 23andMe dataset. The resultant MDD dataset included 135,458 cases and 344,901 healthy controls (18), and the AD dataset included 96,794 cases and 145,775 healthy controls (19). For the inclusion of each dataset, both bi-allelic SNPs and imputation INFO above 0.80 were required. Each SNP was compared between the two datasets, and SNPs with conflicting alleles were excluded. If an SNP was mapped to opposite strands in the two datasets, alleles of this SNP in the second dataset were flipped, and the effect direction was reversed.

Genetic Correlation and Polygenic Overlap Analysis

GWAS summary results were utilized to analyze the genetic correlation of MDD with ADs by LD score regression software (LDSC, v1.0.1) (20, 21). A polygenic overlap was analyzed by MiXeR v1.2 using default parameters (13). Using GWAS summary statistics, MiXeR quantifies the polygenic overlap irrespective of the genetic correlation between traits. Based on the univariate causal mixture model (22), MiXeR builds four bivariate normal distributions, with two causal components for variants specific to each trait, one causal component for variants affecting both traits, and a null component for variants with no effect on either trait. The likelihood function of the observed signed test statistics (GWAS Z-scores) is produced from the prior distribution of genetic effects, incorporating effects of the LD structure, sample size, minor allele frequency (MAF), cryptic relationships, and sample overlap. The summary statistics are used to estimate the parameters of the mixture model by optimization of the likelihood function. The number of causal variants reported by the software is 22.6% of the total estimated variants, which account for 90% of SNP heritability for each trait.

MR Analysis

Bidirectional causal associations between MDD and ADs were inferred using GSMR v1.0.9 (15). Instrumental variants were selected based on default p ≤ 5×10^{-8}. It is well accepted that pleiotropy is a potential source of bias and an inflated estimation in an MR analysis (23). In GSMR, the HEIDI-outlier statistical approach allows the detection and elimination of genetic instruments with apparent pleiotropic effects on both risk factors and disease (15, 24). It was suggested that genetic correlation may confound Mendelian randomization estimates (25). To examine this possibility, we performed a latent causal variable model (LCV) analysis between MDD and ADs (26). The LCV framework utilizes the genetic causality proportion (GCP) to quantify the partial causality of trait 1 on trait 2. The GCP ranges from 0 (no partial genetic causality) to 1 (full genetic causality). A high value of GCP indicates a causal effect of interventions targeting trait 1 on trait 2.

Cross-Trait Meta-Analysis

A cross-trait meta-analysis of the MDD and the ADs was executed by the subset-based fixed-effect method ASSET v2.4.0, which permits the characterization of each SNP with respect to its pattern of effects on multiple phenotypes (27). For each assessed variant, this type of analysis returns a p-value for the best subset containing the studies contributing to the overall association signal. The meta-analysis pools the effect of a given SNP across K studies, weighting the effects by the size of the respective study. After subset-based meta-analysis, SNP-related findings were considered statistically significant, if two-tailed p values were lower than 5×10^{-8}. In the meta-analysis results, functional annotation and gene-mapping of variants and identifying LD-independent genomic regions were performed on a FUMA platform (28). Firstly, independent significant SNPs (IndSigSNPs) were identified based on their p-value being genome-wide significant ($p \leq 5.0 \times 10^{-8}$) and independent of each other ($r^2 < 0.6$). Secondly, lead SNPs were identified as a subset of the independent significant SNPs that were in LD with each other at $r^2 < 0.1$ within a 250-kb window. The gene-based association for the meta-analysis of MDD and ADs was conducted using MAGMA (29).

To ensure that sample overlap did inflate estimates of genetic overlap between MDD and ADs, λmeta statistics, which use effect size concordance to detect sample overlap or heterogeneity, were calculated (30). Under the null hypothesis, λmeta equals 1 when the pair of cohorts are completely independent. When there are overlapping samples, λmeta is less than 1.

Fine-Mapping of TWAS Associations

To prioritize putatively causal genes, fine-mapping of causal gene sets (FOCUS v0.6.10) (31) to the meta-analysis result of MDD and ADs was performed in four relevant tissues, including the brain, whole blood, lung, and skin. Using FOCUS, predicted expression correlations were modeled and posterior inclusion probabilities (PIP) are assigned to genes within each transcriptome-wide association study (TWAS) region in the relevant tissue types. A multi-tissue eQTL reference weight database from the software was used as eQTL weights, while LD information from LDSC was used as a reference. Multiple-testing correction was used to account for all gene–tissue pairs based on Benjamini–Hochberg adjusted TWAS p-values (FDR < 0.05).

Knowledge-Based Analysis

GWAS results, including meta-analysis, were obtained for depression (major depressive disorder and depressive symptoms) and for ADs from the GWAS Catalog database (access date: April 17, 2020) (32). We explore whether the genes shared by MDD and ADs have been identified in previous genome-wide association studies. Protein–protein interaction analysis was conducted using STRING v11 (33). Enrichment of the 27 genes in the GWAS catalog reported genes was analyzed using FUMA (28).

All the statistical analyses were conducted in R 3.6.1 or Python 3.7 environment. A detailed description of the methods is provided in the **Supplementary File**.

RESULTS

Genetic Correlation and Polygenic Overlap Analysis

MDD displayed a significant genetic correlation with ADs (r = 0.18, s.e. = 0.03, p = 1.04×10^{-9}). The LD score intercept did not deviate from zero (0.017). The polygenic analysis highlighted approximately 15.8 thousand variants causally influencing MDD and 0.9 thousand variants influencing ADs. Among these variants, approximately 0.8 thousand variants were shared between the two diseases (**Figure 1A**). MDD has much larger numbers of causal variants than ADs, indicating a higher polygenic property of MDD.

MR Analysis

Mendelian randomization analysis indicated that genetic liability to MDD has a causal effect on ADs (b = 0.22, s.e. = 0.05, OR = 1.25, 95%CI: 1.13–1.37, p = 1.76×10^{-6}), with 45 independent instrumental variants being involved. The genetic liability of ADs conferred a causal effect on MDD (b = 0.05, s.e. = 0.02, OR = 1.05, 95%CI: 1.01–1.09, p = 7.57×10^{-3}), with 115 independent instrumental variants being involved (**Figure 1B**). The LCV analysis showed that GCP was 0.49 (0.32), supporting a causal effect of genetic liability to MDD on ADs.

Cross-Trait Meta-Analysis

The cross-trait meta-analysis of MDD and ADs revealed the involvement of 103 loci, 470 significant independent SNPs (IndSigSNPs), and 141 lead SNPs, including 44 pleiotropic IndSigSNPs located in 18 loci (associated with both traits) (**Figure 2A**, **Table 1** and **Supplementary Tables 1, 2**). The 14q23 locus is shown in **Figure 2B**. A total of 82 pleiotropic protein-coding genes were identified, including 27 protein-coding genes implicated by the pleiotropic IndSigSNPs and another 55 protein-coding genes implicated by SNPs tagged by IndSigSNPs (**Supplementary Table 3**). The gene-based association for the meta-analysis of MDD and ADs identified a total of 273 significant genes at the threshold of 2.70×10^{-6} (Bonferroni correction, 0.05/18,545) (**Supplementary Table 4**). Compared with SNP-based analysis, an additional 63 genes were identified by the gene-based analysis, including DRD2.

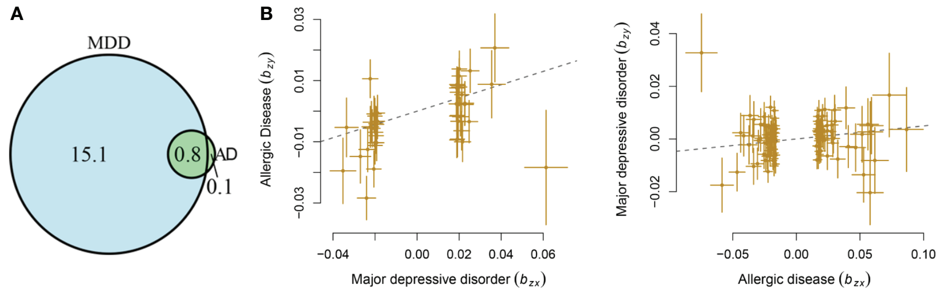


FIGURE 1 | Shared causal variants and causal associations between MDD and ADs. **(A)** Venn diagrams of unique and shared polygenic components at the causal level, showing a polygenic overlap between MDD and ADs. The numbers indicate the estimated quantity of causal variants (in thousands) per component, explaining 90% of SNP heritability in each phenotype. The size of the circles reflects the degree of polygenicity. **(B)** Causal associations between MDD and ADs. The lines denote effect sizes **(B)**. The left panel denotes the causal effect of MDD on ADs. The left panel denotes the causal effect of ADs on MDD.

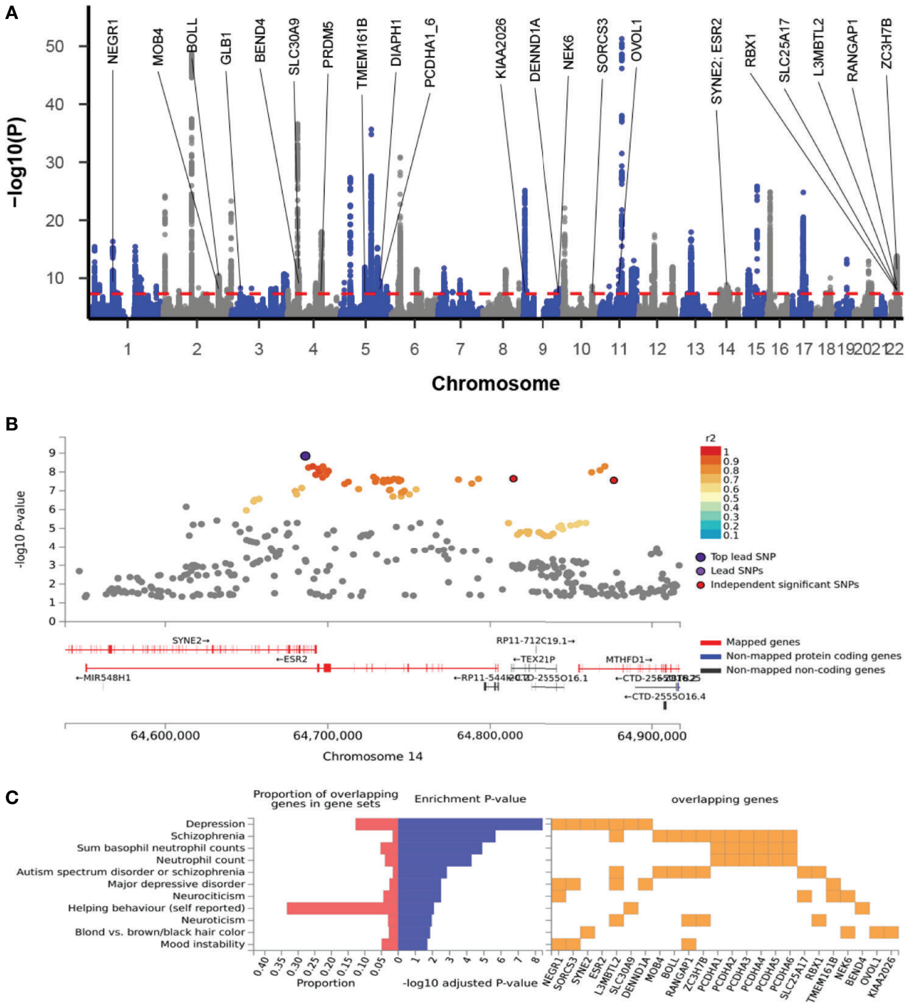


FIGURE 2 | Cross-trait meta-analysis of MDD and ADs. **(A)** Manhattan plot of meta-analysis of MDD with ADs. The x-axis is the chromosomal position of SNPs, and the y-axis is the significance of the SNPs (-$\log_{10}$P). Protein-coding genes containing or adjacent to independent significant SNPs shared by two traits were annotated. PCDHA1_6: *PCDHA1*, *PCDHA2*, *PCDHA3*, *PCDHA4*, *PCDHA5*, and *PCDHA6*. **(B)** The 14q23 locus containing the *ESR2* gene. Each SNP is colored based on the highest r^2 to one of the independent significant SNPs. **(C)** Enrichment of the 27 protein-coding genes in GWAS catalog gene sets.

TABLE 1 | Genomic loci shared between MDD and ADs.

SNP	Chr : BP	P	Start : End	Genes
rs10789340	1:72940273	4.85×10^{-17}	72512988:72958905	**NEGR1**; RPL31P12
rs700646	2:198608511	3.80×10^{-11}	198148191:198954774	**MOB4**; **BOLL**; AC011997.1
rs11927929	3:33087057	5.46×10^{-9}	33068268:33126972	**GLB1**
rs34215985	4:42047778	2.07×10^{-12}	41882601:42187640	RP11-457P14.5; RP11-457P14.6; **SLC30A9**; **BEND4**
rs71600495	4:121628028	1.57×10^{-8}	121625080:121655414	**PRDM5**
rs247910	5:87630769	1.41×10^{-12}	87437079:88065637	**TMEM161B**; TMEM161B-AS1; LINC00461; CTC-467M3.1
rs1363105	5:103917790	1.80×10^{-10}	103671867:104082179	RP11-6N13.1
rs10060640	5:140211226	7.62×10^{-9}	140024042:140222641	**PCDHA1; PCDHA2; PCDHA3; PCDHA4; PCDHA6; PCDHA5**
rs3844598	5:140992235	3.14×10^{-10}	140893490:141032603	**DIAPH1**
rs11135349	5:164523472	2.71×10^{-9}	164465319:164678946	CTC-340A15.2
rs2064219	6:27376001	3.07×10^{-10}	25684606:29607101	MCFD2P1
rs144829310	9:6208030	7.58×10^{-26}	5609742:6621027	AK4P4; **KIAA2026**
rs549779	9:126613028	2.62×10^{-8}	126452936:126714710	**DENND1A**
rs10818936	9:127006346	3.82×10^{-8}	126999153:127144622	**NEK6**
rs61867293	10:106563924	2.60×10^{-9}	106529451:106830537	**SORCS3**
rs479844	11:65551957	3.64×10^{-12}	65401336:65641033	**OVOL1**
rs915057	14:64686207	1.42×10^{-9}	64649894:64877135	**SYNE2; ESR2**
rs136402	22:41598933	1.51×10^{-14}	41085969:42216326	**SLC25A17**; **RBX1**; Y_RNA; RP11-12M9.4; RP1-85F18.5; **L3MBTL2**; **RANGAP1**; **ZC3H7B**

Chr, chromosome; BP, base position. Protein-coding genes are shown in bold.

The λmeta value was at 1.18 for datasets between MDD and ADs, indicating no significant overlap between MDD and AD GWAS samples. Quantile–quantile (QQ) plots to display the observed meta-analysis statistics versus the expected statistics under the null model of no associations in the -log10(p) scale are shown in **Supplementary Figure 1**.

Fine-Mapping of TWAS Associations

To prioritize putatively causal genes, we used the fine-mapping of TWAS associations. A total of 126 gene–tissue pairs were identified between the 82 genes and the four tissues, with 36 genes being associated with two or more tissues (**Supplementary Table 5**). A total of 31 gene–tissue pairs were in the credible sets. Fifteen genes associated with three or more tissues are listed in **Table 2**. However, most genes in **Table 2** had low PIP. Of note, the *ESR2* gene was associated with three tissues (skin, lung, and blood) with relatively high posterior probability (**Figure 3**).

Knowledge-Based Analysis

A total of 23 out of the 27 pleiotropic protein-coding genes have been identified in previous GWASs on depression or ADs (**Supplementary Table 6**). Among these 23 genes were 16 genome-wide risk genes for depression, including *BEND4*, *DENND1A*, *ESR2*, *L3MBTL2*, *NEGR1*, *PCDHA1*, *PCDHA2*, *PCDHA3*, *PCDHA4*, *PCDHA5*, *PCDHA6*, *RBX1*, *SLC30A9*, *SORCS3*, *SYNE2*, and *TMEM161B*, and 8 genome-wide risk genes for ADs, including *BOLL*, *DIAPH1*, *GLB1*, *MOB4*, *NEK6*, *OVOL1*, *RANGAP1*, and *RBX1*. Enrichment of the 27 genes in the GWAS catalog-reported genes revealed that these genes were enriched in several mental disorders and basophil neutrophil counts, as well as neutrophil counts, supporting the involvement of these genes in neurodevelopmental conditions and atopic diseases (**Figure 2C** and **Supplementary Table 7**).

PPI analysis showed that a majority of the 82 genes are interconnected, forming one large network and several small

TABLE 2 | TWAS analysis in the four tissues.

Gene	GWAS P	Chr : Start-End	Tissue	Brain Z (PIP)	Blood Z (PIP)	Lung Z (PIP)	Skin Z (PIP)
SLC30A9	2.07×10^{-12}	4:41992489-42092474	Brain, blood, lung	-6.08 (0.313)	-1.83 (<0.01)	3.72 (<0.01)	
NDUFA2	2.25×10^{-6}	5:140018325-140027370	Brain, blood, skin	6.27 (0.941)	4.28 (0.011)		-2.87 (<0.01)
FCHSD1	5.99×10^{-8}	5:141018869-141030986	Brain, lung, blood	5.64 (0.807)	1.86 (<0.01)	2.42 (<0.01)	
PCDHA7	7.62×10^{-9}	5:140213969-140391929	Lung, brain, skin	-5.02 (0.150)		-5.18 (0.317)	-3.65 (0.013)
WDR55	2.24×10^{-6}	5:140044261-140053709	Skin, blood, lung, brain	-1.56 (<0.01)	-4.65 (0.026)	-1.86 (<0.01)	-4.69 (0.059)
IK	2.25×10^{-6}	5:140026643-140042064	Blood, skin, lung, brain	-3.67 (<0.01)	4.52 (0.034)	-3.7 (<0.01)	-3.37 (<0.01)
TMCO6	2.25×10^{-6}	5:140019012-140024993	Blood, skin, lung, brain	-1.61 (<0.01)	-4.27 (<0.01)	-3.41 (<0.01)	-2.98 (<0.01)
ZMAT2	2.51×10^{-6}	5:140078265-140086248	Lung, blood, skin, brain	-3.5 (<0.01)	2.73 (<0.01)	-2.2 (<0.01)	-2.98 (<0.01)
ZNF391	3.07×10^{-10}	6:27342394-27371683	Brain, skin, blood, lung	2.54 (0.268)	3.83 (<0.01)	-3.86 (<0.01)	3.13 (<0.01)
ESR2	1.42×10^{-9}	14:64550950-64804830	Lung, skin, blood		-5.09 (0.256)	-3.96 (0.243)	-5.58 (0.998)
MTHFD1	5.20×10^{-9}	14:64854749-64926722	Lung, blood, skin		-3.05 (<0.01)	-3.18 (0.018)	-3.32 (<0.01)
MEI1	9.03×10^{-10}	22:42095503-42195460	Brain, skin, lung, blood	6.52 (0.424)	5.88 (<0.01)	5.95 (<0.01)	6.37 (0.042)
XPNPEP3	2.65×10^{-8}	22:41253081-41363838	Blood, skin, lung, brain	3.97 (<0.01)	4.87 (0.045)	5.11 (0.017)	5.42 (0.035)
CCDC134	2.14×10^{-9}	22:42196683-42222303	Brain, lung, blood, skin	3.91 (<0.01)	1.63 (<0.01)	2.49 (<0.01)	-1.82 (<0.01)
DESI1	1.10×10^{-9}	22:41994032-42017100	Lung, skin, brain, blood	2.35 (<0.01)	1.51 (<0.01)	2.34 (<0.01)	2.91 (<0.01)

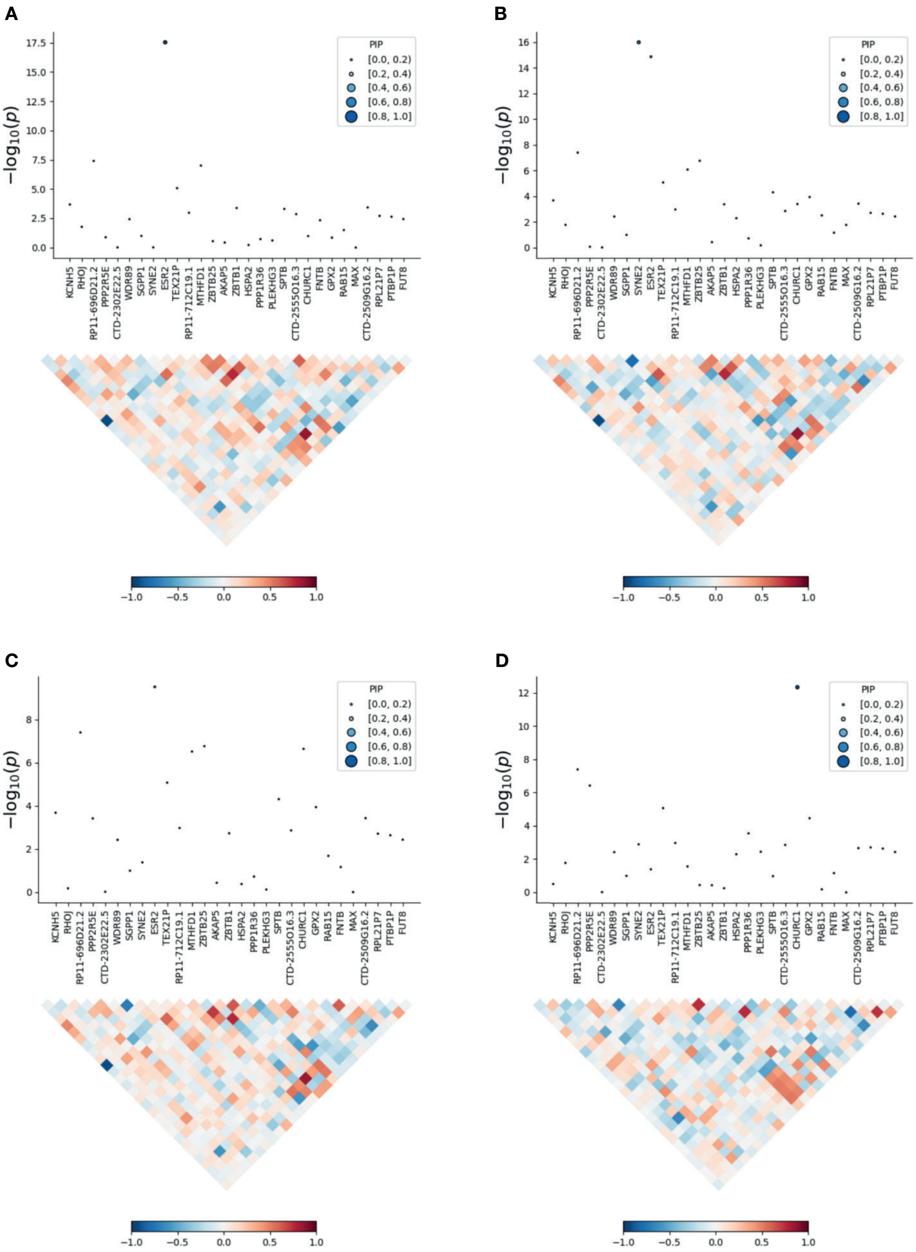


FIGURE 3 | Transcriptome-wide association study of the meta-analysis of MDD and ADs. **(A)** skin; **(B)** blood; **(C)** lung; **(D)** brain. Within each panel, the top part is the transcriptome-wide association signal indicating strength of the predicted expression association with trait, and the bottom part is the induced correlation of the predicted expression.

networks (**Supplementary Figure 2**). Schematics of *ESR2* gene interactions with depression and ADs are shown in **Figure 4**.

DISCUSSION

In this study, we detected a significant genetic correlation between MDD and ADs (r = 0.18), at a level comparable to that for a previously reported correlation of MDD and autism spectrum disorder (r = 0.16) (34). Our results indicate a much higher polygenicity of MDD when compared to ADs, with substantial polygenic overlap between these conditions identified. Nearly 90% of causal variants influencing the risk of ADs may also affect MDD. Cumulative evidence supports a close relationship between these two conditions in the context of underlining genetics.

More importantly, causal relationships between MDD and ADs were discovered. In particular, a major causal effect of

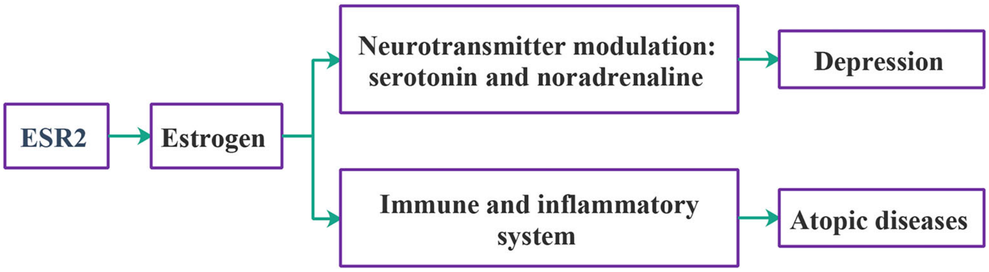

ESR2
Estrogen
Neurotransmitter modulation: serotonin and noradrenaline
Immune and inflammatory system
Depression
Atopic diseases

FIGURE 4 | Schematic relationships of *ESR2* with depression and ADs.

genetic liability to depression on ADs was detected. Although liability to ADs also exerts a statistically significant causal effect on MDD, the size of this effect is relatively small (b = 0.05). Previous studies already showed the possible influence exerted by MDD on ADs. For instance, patients with MDD show elevated levels of non-esterified fatty acids in plasma (35); other studies showed that fatty acids may contribute to the development of atopic diseases such as hay fever and asthma (36). Elevated serum interferon levels may contribute to eczema and also are commonly detected in MDD (37). Moreover, MDD has been shown to stimulate the production of cytokines (38), including IL-13 and IL-6, both of which are also strongly involved in asthma pathogenesis (39). Our findings are consistent with these previous studies and partially explain the previously reported comorbidity of MDD and ADs (8–12), while adding novel insights into underlying pathogenetic mechanisms. Notably, one previous study reported that depression may lead to asthma rather than the opposite (40). The causal effect of ADs on MDD should be further evaluated in additional datasets.

Shared genetic liability between MDD and ADs offers the possibility of employing polygenic risk scores (PRS) for evaluating allergic risks in MDD patients and the risk of developing depression in AD patients. This strategy may lead to an improvement in the clinical management of these conditions. Shared biological markers of MDD and ADs are far from being well studied. The cross-trait analysis revealed that MDD and ADs share 18 loci and a panel of protein-coding genes. The majority of these pleiotropic protein-coding genes have been previously implicated either in depression or in ADs, with a genome-wide significance level. For example, the *RBX1* gene was reported as a significant contributor to both depression (41) and ADs (42). To shed new light on the genetic susceptibility of ADs and MDD, we have concentrated on the estrogen receptor β encoding gene *ESR2* for further discussion.

Estrogen is capable of modulating neurotransmitter turnover to enhance the levels of serotonin and noradrenaline and participates in the regulation of serotonin receptor amounts and function (43). Accumulating evidence indicates the involvement of estrogen signaling in depression (44). In females, estrogen fluctuations are associated with depressed mood (45), and the beneficial effects of estrogen-containing hormone treatments were reported (46, 47). The gene for estrogen receptor β, *ESR2*, has been previously identified as a genome-wide significant gene contributing to MDD (18, 48). As the levels of estrogen are easily modulated by pharmacological means, the association between *ESR2* and MDD may inform the development of personalized treatment modalities for this condition. Notably, model studies in neonatal rats treated with antidepressant clomipramine uncovered both the changes in the levels of estrogen receptors on the surface of brain cells and the neurochemical changes that resemble human depression (49). The role of estrogens in the development of ADs is noticeable as well. Women have a higher prevalence of asthma and display its greater severity than men (50). Estrogen receptors are found on numerous immune-regulatory cells, with estrogen-dependent responses favoring the shift toward allergy. In particular, estrogens promote allergic response by stimulating Th2 polarization, boosting class switching of B cells to IgE production, and prompting mast cell and basophil degranulation (51). *ESR2* and its product, estrogen receptor β, have been suggested as potential targets for asthma treatment (52). There is also accumulating evidence supporting estrogens' role in hay fever and eczema (53, 54). In particular, there is a correlation between the mean number of ER-β-positive cells in the nasal mucosa and seasonal allergy symptoms (55).

This study identified *ESR2* as a novel genome-wide significant contributor to ADs, providing strong support for the involvement of the estrogen pathway in ADs. Fine-mapping of TWAS had assigned the posterior probability for causality for *ESR2* in the skin, blood PBMCs, and lung tissue at 0.998, 0.256, and 0.243, respectively. Although the fine-mapping of TWAS hits did not support the involvement of *ESR2* in the brain, analysis of existing literature points at its role in neurodevelopment and mental disorders. Together, our findings highlight *ESR2* as a critical gene for both MDD and ADs and point to its relevance at the therapy target.

The presented study has several strengths. First, we utilized the largest combination of available datasets as a study backbone. Furthermore, to avoid potential population heterogeneity across the studies, we limited our analysis to individuals of European ancestry. Lastly, the genetic relationship between MDD and ADs was explored systemically by employing multiple analytic frameworks.

However, several limitations should also be noted. The datasets employed in this study only contained data of three subtypes of ADs. Further studies using more datasets covering other subtypes of ADs are warranted to evaluate the associations between MDD and ADs. In TWAS, the gene expression levels are imputed from weighted linear combinations of SNPs and, therefore, may report noise. As our analysis was limited to a genetic component of each trait, hence, the presented results should be interpreted cautiously, with the understanding that human traits arise from a complex web of interactions of various psycho-social-environmental factors.

In summary, our findings reveal shared genetic liability and causal links between MDD and atopic diseases, which may underline the phenotypic relationship between MDD and ADs. Presented results may have implications both for the therapy and for the management of MDD and ADs.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found here: https://www.ebi.ac.uk/gwas.

AUTHOR CONTRIBUTIONS

FZ contributed to the study design and data analysis. HC, SL, and AB contributed to drafting and revising the work. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (81471364).

ACKNOWLEDGMENTS

We thank members of the Psychiatric Genomics Consortium, the UK Biobank, 23andMe, and other teams, who generously shared the GWAS data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.665160/full#supplementary-material

REFERENCES

1. DALYs GBD and Collaborators H. Global, Regional, and National Disability-Adjusted Life-Years (DALYs) for 359 Diseases and Injuries and Healthy Life Expectancy (HALE) for 195 Countries and Territories, 1990-2017: A Systematic Analysis for the Global Burden of Disease Study 2017. *Lancet* (2018) 392(10159):1859–922. doi: 10.1016/S0140-6736(18)32335-3
2. Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJ, et al. Burden of Depressive Disorders by Country, Sex, Age, and Year: Findings From the Global Burden of Disease Study 2010. *PloS Med* (2013) 10(11): e1001547. doi: 10.1371/journal.pmed.1001547
3. Veerman JL, Dowrick C, Ayuso-Mateos JL, Dunn G, Barendregt JJ. Population Prevalence of Depression and Mean Beck Depression Inventory Score. *Br J Psychiatry* (2009) 195(6):516–9. doi: 10.1192/bjp.bp.109.066191
4. Pinart M, Benet M, Annesi-Maesano I, von Berg A, Berdel D, Carlsen KC, et al. Comorbidity of Eczema, Rhinitis, and Asthma in IgE-Sensitised and Non-IgE-Sensitised Children in MeDALL: A Population-Based Cohort Study. *Lancet Respir Med* (2014) 2(2):131–40. doi: 10.1016/S2213-2600(13)70277-7
5. Disease GBD, Injury I, Prevalence C. Global, Regional, and National Incidence, Prevalence, and Years Lived With Disability for 310 Diseases and Injuries, 1990-2015: A Systematic Analysis for the Global Burden of Disease Study 2015. *Lancet* (2016) 388(10053):1545–602. doi: 10.1016/S0140-6736(16)31678-6
6. Mattos JL, Woodard CR, Payne SC. Trends in Common Rhinologic Illnesses: Analysis of U.S. Healthcare Surveys 1995-2007. *Int Forum Allergy Rhinol* (2011) 1(1):3–12. doi: 10.1002/alr.20003
7. Barbarot S, Auziere S, Gadkari A, Girolomoni G, Puig L, Simpson EL, et al. Epidemiology of Atopic Dermatitis in Adults: Results From an International Survey. *Allergy* (2018) 73(6):1284–93. doi: 10.1111/all.13401
8. Misery L, Taieb C, Schollhammer M, Bertolus S, Coulibaly E, Feton-Danou N, et al. Psychological Consequences of the Most Common Dermatoses: Data From the Objectifs Peau Study. *Acta Derm Venereol* (2020) 100(13):adv00175. doi: 10.2340/00015555-3531
9. Choi HG, Kim JH, Park JY, Hwang YI, Jang SH, Jung KS. Association Between Asthma and Depression: A National Cohort Study. *J Allergy Clin Immunol Pract* (2019) 7(4):1239–45.e1231. doi: 10.1016/j.jaip.2018.10.046
10. Schonmann Y, Mansfield KE, Hayes JF, Abuabara K, Roberts A, Smeeth L, et al. Atopic Eczema in Adulthood and Risk of Depression and Anxiety: A Population-Based Cohort Study. *J Allergy Clin Immunol Pract* (2020) 8 (1):248–57.e216. doi: 10.1016/j.jaip.2019.08.030
11. Cheng CM, Hsu JW, Huang KL, Bai YM, Su TP, Li CT, et al. Risk of Developing Major Depressive Disorder and Anxiety Disorders Among Adolescents and Adults With Atopic Dermatitis: A Nationwide Longitudinal Study. *J Affect Disord* (2015) 178:60–5. doi: 10.1016/j.jad.2015.02.025
12. Wan J, Takeshita J, Shin DB, Gelfand JM. Mental Health Impairment Among Children With Atopic Dermatitis: A United States Population-Based Cross-Sectional Study of the 2013-2017 National Health Interview Survey. *J Am Acad Dermatol* (2020) 82(6):1368–75. doi: 10.1016/j.jaad.2019.10.019
13. Frei O, Holland D, Smeland OB, Shadrin AA, Fan CC, Maeland S, et al. Bivariate Causal Mixture Model Quantifies Polygenic Overlap Between Complex Traits Beyond Genetic Correlation. *Nat Commun* (2019) 10 (1):2417. doi: 10.1038/s41467-019-10310-0
14. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian Randomization: Using Genes as Instruments for Making Causal Inferences in Epidemiology. *Stat Med* (2008) 27(8):1133–63. doi: 10.1002/sim.3034
15. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, et al. Causal Associations Between Risk Factors and Common Diseases Inferred From GWAS Summary Data. *Nat Commun* (2018) 9(1):224. doi: 10.1038/s41467-017-02317-2
16. Zhu Z, Zhu X, Liu CL, Shi H, Shen S, Yang Y, et al. Shared Genetics of Asthma and Mental Health Disorders: A Large-Scale Genome-Wide Cross-Trait Analysis. *Eur Respir J* (2019) 54(6):1901507. doi: 10.1183/13993003.01507-2019
17. Zhu Z, Lee PH, Chaffin MD, Chung W, Loh PR, Lu Q, et al. A Genome-Wide Cross-Trait Analysis From UK Biobank Highlights the Shared Genetic Architecture of Asthma and Allergic Diseases. *Nat Genet* (2018) 50(6):857–64. doi: 10.1038/s41588-018-0121-0
18. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-Wide Association Analyses Identify 44 Risk Variants and Refine the Genetic Architecture of Major Depression. *Nat Genet* (2018) 50 (5):668–81. doi: 10.1038/s41588-018-0090-3
19. Ferreira MA, Vonk JM, Baurecht H, Marenholz I, Tian C, Hoffman JD, et al. Shared Genetic Origin of Asthma, Hay Fever and Eczema Elucidates Allergic Disease Biology. *Nat Genet* (2017) 49(12):1752–7. doi: 10.1038/ng.3985
20. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang JSchizophrenia Working Group of the Psychiatric Genomics C, et al. LD Score Regression Distinguishes Confounding From Polygenicity in Genome-Wide Association Studies. *Nat Genet* (2015) 47(3):291–5. doi: 10.1038/ng.3211

21. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An Atlas of Genetic Correlations Across Human Diseases and Traits. *Nat Genet* (2015) 47(11):1236–41. doi: 10.1038/ng.3406
22. Holland D, Frei O, Desikan R, Fan CC, Shadrin AA, Smeland OB, et al. Beyond SNP Heritability: Polygenicity and Discoverability of Phenotypes Estimated With a Univariate Gaussian Mixture Model. *PloS Genet* (2020) 16 (5):e1008612. doi: 10.1371/journal.pgen.1008612
23. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in Complex Traits: Challenges and Strategies. *Nat Rev Genet* (2013) 14(7):483–95. doi: 10.1038/nrg3461
24. Ong JS, MacGregor S. Implementing MR-PRESSO and GCTA-GSMR for Pleiotropy Assessment in Mendelian Randomization Studies From a Practitioner's Perspective. *Genet Epidemiol* (2019) 43(6):609–16. doi: 10.1002/gepi.22207
25. Verbanck M, Chen CY, Neale B, Do R. Detection of Widespread Horizontal Pleiotropy in Causal Relationships Inferred From Mendelian Randomization Between Complex Traits and Diseases. *Nat Genet* (2018) 50(5):693–8. doi: 10.1038/s41588-018-0099-7
26. O'Connor LJ, Price AL. Distinguishing Genetic Correlation From Causation Across 52 Diseases and Complex Traits. *Nat Genet* (2018) 50(12):1728–34. doi: 10.1038/s41588-018-0255-0
27. Bhattacharjee S, Rajaraman P, Jacobs KB, Wheeler WA, Melin BS, Hartge P, et al. A Subset-Based Approach Improves Power and Interpretation for the Combined Analysis of Genetic Association Studies of Heterogeneous Traits. *Am J Hum Genet* (2012) 90(5):821–35. doi: 10.1016/j.ajhg.2012.03.015
28. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional Mapping and Annotation of Genetic Associations With FUMA. *Nat Commun* (2017) 8 (1):1826. doi: 10.1038/s41467-017-01261-5
29. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PloS Comput Biol* (2015) 11(4):e1004219. doi: 10.1371/journal.pcbi.1004219
30. Chen GB, Lee SH, Zhu ZX, Benyamin B, Robinson MR. EigenGWAS: Finding Loci Under Selection Through Genome-Wide Association Studies of Eigenvectors in Structured Populations. *Hered (Edinb)* (2016) 117(1):51–61. doi: 10.1038/hdy.2016.25
31. Mancuso N, Freund MK, Johnson R, Shi H, Kichaev G, Gusev A, et al. Probabilistic Fine-Mapping of Transcriptome-Wide Association Studies. *Nat Genet* (2019) 51(4):675–82. doi: 10.1038/s41588-019-0367-1
32. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of Published Genome-Wide Association Studies, Targeted Arrays and Summary Statistics 2019. *Nucleic Acids Res* (2019) 47(D1):D1005–12. doi: 10.1093/nar/gky1120
33. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING V11: Protein-Protein Association Networks With Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. *Nucleic Acids Res* (2019) 47(D1):D607–13. doi: 10.1093/nar/gky1131
34. Brainstorm C, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. Analysis of Shared Heritability in Common Disorders of the Brain. *Science* (2018) 360(6395):eaap8757. doi: 10.1126/science.aap8757
35. Badawy AA. Tryptophan: The Key to Boosting Brain Serotonin Synthesis in Depressive Illness. *J Psychopharmacol* (2013) 27(10):878–93. doi: 10.1177/0269881113499209
36. Kompauer I, Demmelmair H, Koletzko B, Bolte G, Linseisen J, Heinrich J. Association of Fatty Acids in Serum Phospholipids With Hay Fever, Specific and Total Immunoglobulin E. *Br J Nutr* (2005) 93(4):529–35. doi: 10.1079/bjn20041387
37. Berger L, Descamps V, Marck Y, Dehen L, Grossin M, Crickx B, et al. Alpha Interferon-Induced Eczema in Atopic Patients Infected by Hepatitis C Virus: 4 Case Reports. *Ann Dermatol Venereol* (2000) 127(1):51–5.
38. Wright CE, Strike PC, Brydon L, Steptoe A. Acute Inflammation and Negative Mood: Mediation by Cytokine Activation. *Brain Behav Immun* (2005) 19 (4):345–50. doi: 10.1016/j.bbi.2004.10.003
39. Bullone M, Lavoie JP. The Equine Asthma Model of Airway Remodeling: From a Veterinary to a Human Perspective. *Cell Tissue Res* (2020) 380(2):223–36. doi: 10.1007/s00441-019-03117-4
40. Gao YH, Zhao HS, Zhang FR, Gao Y, Shen P, Chen RC, et al. The Relationship Between Depression and Asthma: A Meta-Analysis of Prospective Studies. *PloS One* (2015) 10(7):e0132424. doi: 10.1371/journal.pone.0132424
41. Nagel M, Jansen PR, Stringer S, Watanabe K, de Leeuw CA, Bryois J, et al. Meta-Analysis of Genome-Wide Association Studies for Neuroticism in 449,484 Individuals Identifies Novel Genetic Loci and Pathways. *Nat Genet* (2018) 50(7):920–7. doi: 10.1038/s41588-018-0151-7
42. Johansson A, Rask-Andersen M, Karlsson T, Ek WE. Genome-Wide Association Analysis of 350 000 Caucasians From the UK Biobank Identifies Novel Loci for Asthma, Hay Fever and Eczema. *Hum Mol Genet* (2019) 28(23):4022–41. doi: 10.1093/hmg/ddz175
43. Summer BE, Fink G. Estrogen Increases the Density of 5-Hydroxytryptamine(2A) Receptors in Cerebral Cortex and Nucleus Accumbens in the Female Rat. *J Steroid Biochem Mol Biol* (1995) 54(1-2):15–20. doi: 10.1016/0960-0760(95)00075-b
44. Ancelin ML, Scali J, Ritchie K. Hormonal Therapy and Depression: Are We Overlooking an Important Therapeutic Alternative? *J Psychosom Res* (2007) 62(4):473–85. doi: 10.1016/j.jpsychores.2006.12.019
45. Soares CN, Zitek B. Reproductive Hormone Sensitivity and Risk for Depression Across the Female Life Cycle: A Continuum of Vulnerability? *J Psychiatry Neurosci* (2008) 33(4):331–43.
46. Zweifel JE, O'Brien WH. A Meta-Analysis of the Effect of Hormone Replacement Therapy Upon Depressed Mood. *Psychoneuroendocrinology* (1997) 22(3):189–212. doi: 10.1016/s0306-4530(96)00034-0
47. Gleason CE, Dowling NM, Wharton W, Manson JE, Miller VM, Atwood CS, et al. Effects of Hormone Therapy on Cognition and Mood in Recently Postmenopausal Women: Findings From the Randomized, Controlled KEEPS-Cognitive and Affective Study. *PloS Med* (2015) 12(6):e1001833. doi: 10.1371/journal.pmed.1001833. discussion e1001833.
48. Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shirali M, et al. Genome-Wide Meta-Analysis of Depression Identifies 102 Independent Variants and Highlights the Importance of the Prefrontal Brain Regions. *Nat Neurosci* (2019) 22(3):343–52. doi: 10.1038/s41593-018-0326-7
49. Limon-Morales O, Arteaga-Silva M, Rojas-Castaneda JC, Molina-Jimenez T, Guadarrama-Cruz GV, Cerbon M, et al. Neonatal Treatment With Clomipramine Modifies the Expression of Estrogen Receptors in Brain Areas of Male Adult Rats. *Brain Res* (2019) 1724:146443. doi: 10.1016/j.brainres.2019.146443
50. Yung JA, Fuseini H, Newcomb DC. Hormones, Sex, and Asthma. *Ann Allergy Asthma Immunol* (2018) 120(5):488–94. doi: 10.1016/j.anai.2018.01.016
51. Fan Z, Che H, Yang S, Chen C. Estrogen and Estrogen Receptor Signaling Promotes Allergic Immune Responses: Effects on Immune Cells, Cytokines, and Inflammatory Factors Involved in Allergy. *Allergol Immunopathol (Madr)* (2019) 47(5):506–12. doi: 10.1016/j.aller.2019.03.001
52. Wang Y, Chen YJ, Xiang C, Jiang GW, Xu YD, Yin LM, et al. Discovery of Potential Asthma Targets Based on the Clinical Efficacy of Traditional Chinese Medicine Formulas. *J Ethnopharmacol* (2020) 252:112635. doi: 10.1016/j.jep.2020.112635
53. Kanda N, Hoashi T, Saeki H. The Roles of Sex Hormones in the Course of Atopic Dermatitis. *Int J Mol Sci* (2019) 20(19):4660. doi: 10.3390/ijms20194660
54. Klis K, Wronka I. Association of Estrogen-Related Traits With Allergic Rhinitis. *Adv Exp Med Biol* (2017) 968:71–8. doi: 10.1007/5584_2016_190
55. Philpott CM, Wild DC, Wolstensholme CR, Murty GE. The Presence of Ovarian Hormone Receptors in the Nasal Mucosa and Their Relationship to Nasal Symptoms. *Rhinology* (2008) 46(3):221–5.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers
in Immunology

ORIGINAL RESEARCH
published: 12 October 2021
doi: 10.3389/fimmu.2021.729501

Influenza Vaccination Is Associated With Lower Incidental Asthma Risk in Patients With Atopic Dermatitis: A Nationwide Cohort Study

Kun Hong Li[1,2], *Pui-Ying Leong*[1,3,4], *Chung-Fang Tseng*[1], *Yu Hsun Wang*[5] *and James Cheng-Chung Wei*[1,4,6]*

[1] Institute of Medicine, Chung Shan Medical University, Taichung, Taiwan, [2] Department of Family Medicine, Changhua Christian Hospital, Changhua, Taiwan, [3] Department of Medicine, Chung Shan Medical University Hospital, Taichung, Taiwan, [4] Division of Allergy, Immunology and Rheumatology, Chung Shan Medical University Hospital, Taichung, Taiwan, [5] Department of Medical Research, Chung Shan Medical University Hospital, Taichung, Taiwan, [6] Graduate Institute of Integrated Medicine, China Medical University, Taichung, Taiwan

OPEN ACCESS

Edited by:
Xu-jie Zhou,
Peking University First Hospital, China

Reviewed by:
Cheryl Rockwell,
Michigan State University,
United States
Jarmila Celakovska,
Charles University, Czechia

Correspondence:
James Cheng-Chung Wei
jccwei@gmail.com

Specialty section:
This article was submitted to Autoimmune and Autoinflammatory Disorders, a section of the journal Frontiers in Immunology

Received: *23 June 2021*
Accepted: *24 September 2021*
Published: *12 October 2021*

Citation:
Li KH, Leong P-Y, Tseng C-F, Wang YH and Wei JC-C (2021) Influenza Vaccination Is Associated With Lower Incidental Asthma Risk in Patients With Atopic Dermatitis: A Nationwide Cohort Study. Front. Immunol. 12:729501. doi: 10.3389/fimmu.2021.729501

Background: Atopic march refers to the natural history of atopic dermatitis (AD) in infancy followed by subsequent allergic rhinitis and asthma in later life. Respiratory viruses interact with allergic sensitization to promote recurrent wheezing and the development of asthma. We aimed to evaluate whether influenza vaccination reduces asthma risk in people with AD.

Methods: This cohort study was conducted retrospectively from 2000 to 2013 by the National Health Insurance Research Database (NHIRD). Patients with newly diagnosed AD (International Classification of Diseases, Ninth Revision, Clinical Modification code 691) were enrolled as the AD cohort. We matched each vaccinated patient with one non-vaccinated patient according to age and sex. We observed each participant until their first asthma event, or the end of the study on December 31, 2013, whichever came first.

Results: Our analyses included 4,414 people with a mean age of 53 years. Of these, 43.8 were male. The incidence density of asthma was 12.6 per 1,000 person-years for vaccinated patients, and 15.1 per 1000 person-years for non-vaccinated patients. The adjusted hazard ratio (aHR) of asthma in the vaccinated cohort relative to the non-vaccinated cohort was 0.69 (95% CI = 0.55–0.87). Vaccinated patients had a lower cumulative incidence of asthma than unvaccinated patients. Vaccinated participants in all age and sex groups trended toward a lower risk of asthma. People will reduce more asthma risk when taking shots every year.

Conclusion: Influenza vaccination was associated with lower asthma risk in patients with AD.

Keywords: asthma, atopic dermatitis, big data analysis, influenza vaccination, Taiwan national health insurance research database

INTRODUCTION

Up to 80% of children with atopic dermatitis (AD) develop asthma or allergic rhinitis later in life (1). Teenagers with asthma have higher AD rates than those without asthma (risk ratio 4.5, 95% CI = 3.1–6.5) (2). The most common mechanisms for this occurrence are barrier defects, skin inflammation, and microbiome alterations that trigger the T-helper type 2 (Th2) cell response and lead to hypersensitization for later disorders. AD and asthma share similar genetic loci, and people afflicted by these conditions often share similar food allergies and early environmental exposures. Vitamin D, probiotics, allergen avoidance, allergen immunotherapy, IgE antagonists, and respiratory infection prevention are all considered strategies for asthma prevention (3). Viral respiratory tract infections in infancy, particularly influenza virus (4–6), respiratory syncytial virus, and human rhinovirus may predict asthma development from late childhood through early adulthood (7). Respiratory viruses interact with allergic sensitization and other microbes to promote recurrent virus- induced wheezing and asthma development *via* a number of mechanisms, including increased inflammatory cell recruitment, promotion of cytokine production, enhanced allergic inflammation, and augmented airway hyperresponsiveness (8).

Children and adults with asthma have increased risks of hospitalization and respiratory morbidity due to acute influenza respiratory infections (9). Children with asthma were particularly prone to increased intensive care unit (ICU) stays and pneumonia during the 2009 H1N1 influenza pandemic. Influenza is also associated with a higher risk of emergency department (ED) treatment failure for acute respiratory illness (10).

To protect asthmatic patients, annual influenza vaccination is recommended by the Advisory Committee for Immunization Practices (ACIP), the American Academy of Pediatrics (AAP), and the Expert Panel for the Diagnosis and Management of Asthma. A recent systemic review and meta- analysis including observational studies suggested that influenza vaccination reduced the risk of asthma's exacerbation (11).

However, the association between influenza vaccination and further asthma risk has not yet been explored, particularly in patients with AD. We conducted an original longitudinal nationwide cohort study to determine whether influenza vaccination in AD patients decreases the risk of asthma.

MATERIALS AND METHODS

Data Source

We analyzed anonymous data from the Taiwan National Health Insurance Research Database (NHIRD). The Taiwan National Health Insurance (NHI) program launched in 1995 and covered 99% of Taiwanese residents. The comprehensive claims data from the NHI program were collected into NHIRD. We utilized the outpatient and inpatient records of one million randomly sampled people in the Longitudinal Health Insurance Database. There were no statistically significant differences in sex, age, or healthcare costs between the sample group and all enrollees. The database used the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes to define the patients' corresponding diseases. Due to safety and privacy concerns, all patient identification information was encrypted.

Study Population

This cohort study was conducted retrospectively from 2000 to 2013. Newly diagnosed AD patients (ICD-9-CM = 691) with at least three outpatient visits or one inpatient admission record were enrolled as the AD cohort. The inclusion criteria were defined as people who received an influenza vaccine after AD diagnosis and before January 1, 2013. The earliest influenza vaccination date was used as the index date for the vaccinated group. We excluded people with a history of asthma (ICD-9-CM = 493) prior to the index date. Patients who never received an influenza vaccination after AD diagnosis were enrolled as the non-vaccinated cohort. The index date of the unvaccinated group was assigned by 1:1 age and sex matching. We observed each participant until the first asthma event or the end of the study on December 31, 2013, whichever came first. This study was approved by the NHIRD research committee and the Joint Institutional Review Board of Chung Shan Medical University (IRB permit number: CS19009).

Main Explanatory Variable

The main explanatory variable in this study was influenza vaccination, which was identified using the therapeutic treatment code A2001C, the ICD-9-CM codes V04.7 or V04.8, or the medication codes for influenza vaccination. The flu usually strikes between November and the following March. In Taiwan, influenza vaccination has been free for people with high-risk comorbidities since 2001, and for all adults over 65 since 1998. This was extended to infants and children between the ages of 6 months and 2 years in 2004, with gradual extension to fifth- and sixth-grade–aged elementary school students in 2012. The vaccines in the influenza immunization program were Adimflu-S (ADIMMUNE Corporation, Taiwan); Fluvirin (Novartis Vaccines, Switzerland); AGRIPPAL S1 (Novartis Vaccines, Switzerland); Begrivac (Novartis Vaccines, Switzerland); Vaxigrip (Pasteur Merieux Connaught, France); and Fluarix (GlaxoSmithKline, USA). Each influenza vaccination program recorded the influenza vaccination status of all vaccinated participants.

Main Outcome and Comorbidities

Asthma was the primary endpoint in this study. We defined asthma patients as patients with three or more outpatient visits or one inpatient record of ICD-9-CM code 493. The following comorbidities developed before the index date were considered as confounders (ICD-9-CM codes indicated in parentheses): hypertension (401–405); hyperlipidemia (272.0–272.4); chronic liver disease (571); chronic kidney disease (585); diabetes (250); gastroesophageal reflux disease (530.11, 530.81, 530.85); allergic rhinitis (472, 473, 477); urticaria (708); chronic obstructive pulmonary disease (COPD; 491, 492, 496); obstructive sleep apnea (327.23, 780.51, 780.52, 780.53, 780.57); cellulitis (682); attention deficit hyperactivity disorder (314.0); anxiety (300.0); and depression (296.2, 296.3, 300.4, 309.0, 309.1, 309.28, 311).

Statistical Analysis

The data were computed as percentages for categorical variables, and as means for continuous variables. We examined the baseline variables of the vaccinated and non-vaccinated cohorts using chi-square, Fisher's exact, and Student's t-tests. A Kaplan–Meier analysis was used to acquire the cumulative incidence curve of each cohort.

We applied the Cox proportional hazards model to estimate asthma's hazard ratio. We also performed a subgroup analysis of the association between influenza vaccination and asthma development. Finally, we performed a sensitivity test to explore the effect of time on this study. All statistical analyses were conducted with SPSS version 18.0 (SPSS Inc., Chicago, IL, USA). P-values of less than 0.05 were considered significant.

RESULTS

This study identified 31,134 people in Taiwan who had newly diagnosed AD (**Figure 1**). During the follow-up period, 20.6% (6,416 people) received an influenza vaccination. After excluding 1,887 people with antecedent asthma, we included the remaining 2,207 people with and without vaccination in each group after sex and age matching in our analyses (**Table 1**).

The majority (74%) of our study's participants were over 40 years old. The mean age was 53 years old. Of the total number of participants, 43.8% were male and 56.2% were female. Vaccinated participants had a slightly higher proportion of concurrent conditions than did unvaccinated participants, including hypertension, chronic kidney disease, diabetes, COPD, and depression.

The asthma incidence density of vaccinated participants was 12.6 per 1,000 person-years, whereas that of non-vaccinated participants was 15.1 per 1,000 person-years (**Table 2**). The adjusted hazard ratio (aHR) of asthma for the vaccinated cohort relative to the non-vaccinated cohort was 0.69 (95% CI = 0.55–0.87). Allergic rhinitis and COPD increased the risk of asthma in participants by 2.37-fold (95% CI = 1.70–3.31) and 2.53-fold (95% CI = 1.62–3.94), respectively.

The cumulative incidence of asthma was lower in vaccinated participants than in non-vaccinated participants (**Figure 2**). However, this difference was not statistically significant (log-rank test p = 0.072).

Table 3 presents the stratification analysis of the association between influenza vaccination and asthma by age and sex. Vaccinated participants in all age groups trended toward a lower risk of asthma (hazard ratio (HR) = 0.54–0.94), though this trend was not statistically significant. Influenza vaccination did not significantly reduce the risk of asthma in men or women; this was attributed to the small sample size. However, a decreasing trend was identified clinically in both men and women, with HR of 0.76 and 0.90, respectively.

We conducted sensitivity analyses to see if the associations of influenza vaccination and incident asthma affected by different follow up duration (**Table 4**). Influenza vaccination significantly and consistently reduced the risk of incident asthma in the last three to seven years of the study.

Table 5 presented the asthma risk reduction is dose dependent. Compared with non-vaccinated group, people who received once flu vaccination didn't have great effect on asthma reduction. But, for those more than 2 times, the HR was 0.59

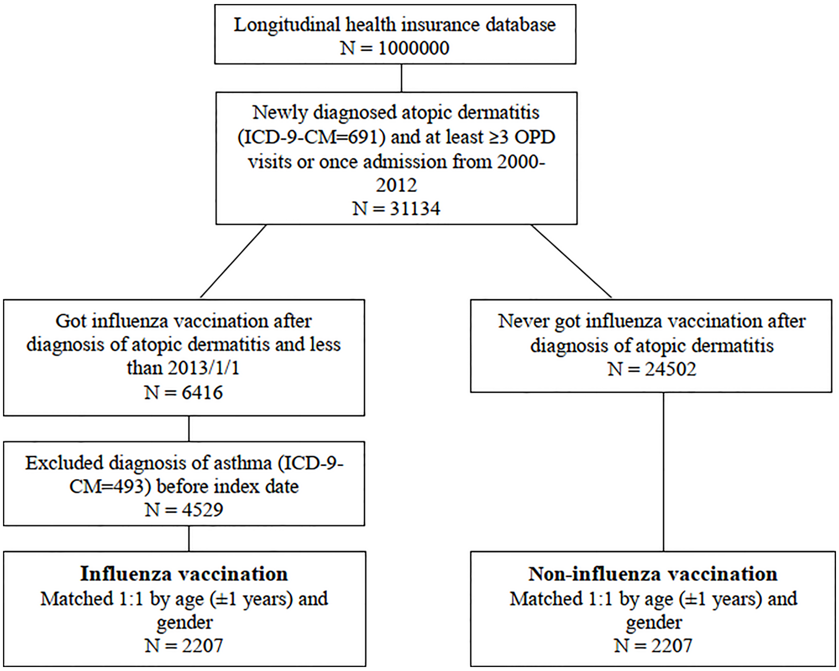

Longitudinal health insurance database
N = 1000000
Newly diagnosed atopic dermatitis (ICD-9-CM=691) and at least ≥3 OPD visits or once admission from 2000-2012
N = 31134
Got influenza vaccination after diagnosis of atopic dermatitis and less than 2013/1/1
N = 6416
Never got influenza vaccination after diagnosis of atopic dermatitis
N = 24502
Excluded diagnosis of asthma (ICD-9-CM=493) before index date
N = 4529
Influenza vaccination
Matched 1:1 by age (±1 years) and gender
N = 2207
Non-influenza vaccination
Matched 1:1 by age (±1 years) and gender
N = 2207

FIGURE 1 | Flowchart of subject's enrollment with and without influenza vaccination.

TABLE 1 | Demographic characteristics of influenza vaccination group and non-influenza vaccination.

	Influenza vaccination (N = 2207)	Non-influenza vaccination (N = 2207)	p value
Age			<0.001
<20	146 (6.6)	151 (6.8)	
20-39	411 (18.6)	418 (18.9)	
40-64	752 (34.1)	948 (43.0)	
≥65	898 (40.7)	690 (31.3)	
Mean ± SD	53.5 ± 19.1	53.1 ± 18.8	0.404
Gender			1.000
Female	1240 (56.2)	1240 (56.2)	
Male	967 (43.8)	967 (43.8)	
Hypertension	647 (29.3)	515 (23.3)	<0.001
Hyperlipidemia	235 (10.6)	194 (8.8)	0.037
Chronic liver disease	118 (5.3)	84 (3.8)	0.014
Chronic kidney disease	45 (2.0)	12 (0.5)	<0.001
Diabetes	364 (16.5)	263 (11.9)	<0.001
GERD	49 (2.2)	34 (1.5)	0.096
Allergic rhinitis	166 (7.5)	144 (6.5)	0.195
Urticaria	139 (6.3)	97 (4.4)	0.005
COPD	80 (3.6)	40 (1.8)	<0.001
OSA	100 (4.5)	79 (3.6)	0.109
Cellulitis	68 (3.1)	44 (2.0)	0.022
ADHD	0 (0.0)	2 (0.1)	0.500[†]
Anxiety	107 (4.8)	105 (4.8)	0.888
Depression	96 (4.3)	54 (2.4)	<0.001

[†] *Fisher's exact test*

COPD, Chronic obstructive pulmonary disease.

OSA, Obstructive sleep apnea.

ADHD, attention-deficit hyperactivity disorder.

TABLE 2 | Cox proportional hazard model analysis for risk of asthma.

	No. of asthma	PY	ID	Univariate		Multivariate	
				HR (95% C.I.)	p value	HR (95% C.I.)	p value
Group							
Non-influenza vaccination	172	11406	15.1	1		1	
Influenza vaccination	134	10639	12.6	0.81 (0.65-1.02)	0.073	0.69 (0.55-0.87)	0.002
Age							
<20	20	1463	13.7	1		1	
20-39	40	3645	11.0	0.76 (0.45-1.30)	0.323	0.77 (0.45-1.33)	0.351
40-64	92	8259	11.1	0.80 (0.49-1.30)	0.371	0.75 (0.46-1.24)	0.263
≥65	154	8677	17.7	1.34 (0.84-2.14)	0.222	1.26 (0.77-2.06)	0.360
Gender							
Female	190	12221	15.5	1		1	
Male	116	9824	11.8	0.77 (0.61-0.97)	0.024	0.75 (0.59-0.95)	0.015
Hypertension	101	5606	18.0	1.42 (1.12-1.81)	0.004	1.15 (0.87-1.51)	0.321
Hyperlipidemia	29	1903	15.2	1.06 (0.72-1.56)	0.759	0.87 (0.58-1.30)	0.495
Chronic liver disease	11	967	11.4	0.80 (0.44-1.45)	0.458	0.72 (0.39-1.32)	0.283
Chronic kidney disease	2	214	9.3	0.60 (0.15-2.43)	0.477	0.61 (0.15-2.48)	0.490
Diabetes	51	3012	16.9	1.25 (0.92-1.69)	0.148	1.12 (0.81-1.56)	0.494
GERD	9	269	33.4	2.13 (1.09-4.13)	0.026	1.78 (0.91-3.50)	0.095
Allergic rhinitis	45	1323	34.0	2.59 (1.89-3.56)	<0.001	2.37 (1.70-3.31)	<0.001
Urticaria	25	1082	23.1	1.68 (1.12-2.53)	0.013	1.51 (1.00-2.29)	0.052
COPD	24	480	50.0	3.64 (2.40-5.52)	<0.001	2.53 (1.62-3.94)	<0.001
OSA	21	801	26.2	1.87 (1.20-2.92)	0.006	1.34 (0.83-2.18)	0.233
Cellulitis	8	557	14.4	1.02 (0.51-2.07)	0.947	0.86 (0.42-1.74)	0.676
Anxiety	27	989	27.3	2.03 (1.36-3.01)	<0.001	1.46 (0.94-2.26)	0.096
Depression	17	624	27.3	1.92 (1.18-3.13)	0.009	1.36 (0.78-2.36)	0.273

ID, Incidence density (per 1000 person-years).

PY, person-years.

COPD, Chronic obstructive pulmonary disease.

OSA, Obstructive sleep apnea.

Multivariate, adjusted for age, gender, hypertension, hyperlipidemia, chronic liver disease, chronic kidney disease, diabetes, GERD, allergic rhinitis, urticaria, COPD, OSA, cellulitis, anxiety, and depression.

(0.44-0.79). We inferred that people will reduce more asthma risk when taking shots every year.

DISCUSSION

Principal Findings

Our 13-year retrospective cohort study revealed a decreased risk of asthma in vaccinated AD patients with an aHR of 0.69 (95% CI = 0.55–0.87) after adjusting for the age, sex, and confounders listed in **Table 2**. Although our subgroup analysis did not yield significant results, we found that influenza vaccination reduced asthma development in all age and sex groups.

Theoretical Mechanism

Atopic march, sometimes called allergic march, refers to AD's natural history and typical progression in infancy, followed by subsequent allergic rhinitis and asthma in later childhood. We have analyzed Influenza infection times between vaccinated and non-vaccinated group. There are no significant infection times differences with and without Influenza vaccination throughout the study period. The mechanism may not be that influenza vaccination directly prevents and/or mitigates the severity of influenza infection, but immune modulation is favor. Allergic sensitization and bacterial colonization due to a dysfunctional skin barrier promote Th2 immunity, which induces systemic responses in the respiratory tract (12). On the other hand, the acute stage of influenza not only causes inflammation and tissue damage to the respiratory tract, but also enhances unrelated local allergic responses *via* the Th2 response (13). Accordingly, the Th2 subset may not play a primary role in virus clearance and recovery, and may lead to immune-mediated injury potentiation (14–16). One way to prevent subsequent asthma and atopic disorders is by using vaccination to restore the Th1/Th2 balance in favor of Th1. An animal study by Skevaki found that influenza-infected animals showed heterologous immunity toward allergens. Immunization *via* vaccination with influenza-derived peptides provided asthma protection through the interferon-gamma response (17). Another study showed influenza vaccination to activate CD4+ and Th1–type cells, which induced the secretion of Th1-type cytokines and promoted T cell immunity (18).

Clinical Implications

An estimated 65%–80% of children with AD develop symptoms in their first year of life. Asthma has a later onset, occurring in only 42% of children in their first year. However, 92% of affected individuals develop symptoms by age 8 (19). Furthermore, patients who experienced asthma onset prior to 1 year of age were reported to have more severe symptoms and greater medical expenses than patients who developed asthma symptoms between 5 and 9 years of age (20). Therefore, efforts to delay early asthma onset will be essential to prevent atopic march and the subsequent outcome of asthma. As shown in the illustration in **Figure 2** and **Table 4**, influenza vaccination can delay the onset of asthma by at least seven years. According to the 2006 report by the International Study of Asthma and Allergies in Childhood (21), AD has a prevalence of 6.7% in 6-to-7-year-old Taiwanese children, but only 4.1% in 13-to-14-year-old Taiwanese children. Such findings indicate that age-related physiological changes may lessen the symptoms of AD or cause them to disappear spontaneously.

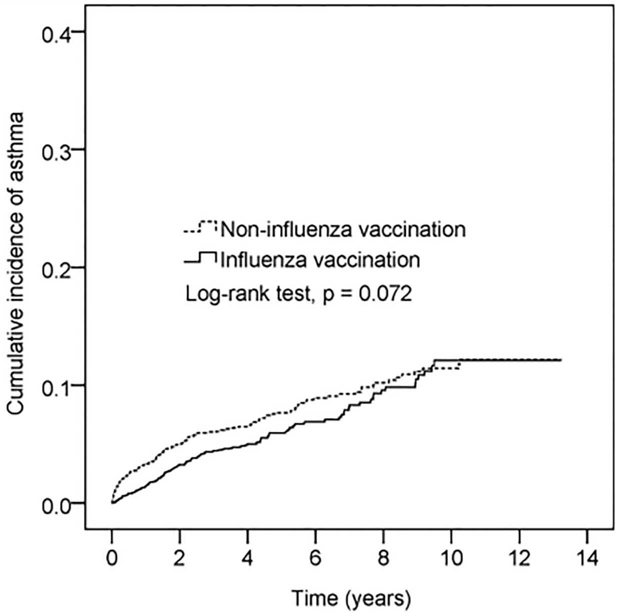


FIGURE 2 | Cumulative incidence of asthma in patients with vaccination *vs* without vaccination. The x-axis represents years after flu vaccination. The index date of the unvaccinated group was assigned by 1:1 age and sex matching. Y-axis represents cumulative incidence of asthma.

TABLE 3 | Subgroup analysis of the association between influenza vaccination and asthma development.

	Influenza vaccination		Non-influenza vaccination		HR (95% C.I.)	p value
	N	No. of asthma	N	No. of asthma		
Age						
<20	146	7	151	13	0.54 (0.22-1.36)	0.190
20-39	411	15	418	25	0.60 (0.32-1.14)	0.117
40-64	752	32	948	60	0.69 (0.45-1.06)	0.092
≥65	898	80	690	74	0.94 (0.68-1.29)	0.685
p for interaction = 0.434						
Gender						
Female	1240	81	1240	109	0.76 (0.57-1.01)	0.057
Male	967	53	967	63	0.90 (0.63-1.30)	0.588
p for interaction = 0.433						

TABLE 4 | Sensitivity analysis of the association between influenza vaccination and asthma development.

	N	No. of asthma	Univariate		Multivariate	
			HR (95% C.I.)	p value	HR (95% C.I.)	p value
Follow-up duration ≤3 years						
Group						
Non-influenza vaccination	2207	129	1		1	
Influenza vaccination	2207	92	0.71 (0.55-0.93)	0.013	0.61 (0.47-0.81)	<0.001
Follow-up duration ≤5 years						
Group						
Non-influenza vaccination	2207	149	1		1	
Influenza vaccination	2207	109	0.74 (0.58-0.95)	0.016	0.63 (0.49-0.81)	<0.001
Follow-up duration ≤7 years						
Group						
Non-influenza vaccination	2207	160	1		1	
Influenza vaccination	2207	122	0.78 (0.62-0.99)	0.040	0.66 (0.52-0.84)	<0.001

Multivariate: adjusted for age, gender, hypertension, hyperlipidemia, chronic liver disease, chronic kidney disease, diabetes, GERD, allergic rhinitis, urticaria, COPD, OSA, cellulitis, anxiety, and depression.

TABLE 5 | influenza vaccination dose response on asthma risk reduction.

	Number	No. of asthma	HR[†] (95% C.I.)	p value
Group				
Non-influenza vaccination	2207	172	1	
Influenza vaccination =1 time	1151	63	0.91 (0.67-1.23)	0.535
Influenza vaccination ≥2 times	1056	71	0.59 (0.44-0.79)	<0.001

[†] *Adjusted for age, gender, hypertension, hyperlipidemia, chronic liver disease, chronic kidney disease, diabetes, GERD, allergic rhinitis, urticaria, COPD, OSA, cellulitis, anxiety, and depression.*

This study evaluated patients who received free influenza vaccinations that were provided in Taiwan to children, the elderly, and people with comorbidities. We found that the vaccinated group had an increased percentage of underlying chronic disease than did the non-vaccinated group. Because asthma is associated with hypertension, diabetes mellitus, dyslipidemia, and cardiovascular disease (22), this difference underestimates the protective effect of influenza vaccination on asthma development. Furthermore, in our stratification study investigating the association of asthma with other comorbidities, we found that asthma increased the risk of developing allergic rhinitis and COPD by 2.37-fold and 2.53-fold, respectively. Therefore, we conclude that people with allergic diseases may benefit from influenza vaccination.

Strengths and Limitations

The primary strength of this study is that it included a long-term comprehensive follow-up from 2000–2013, with universal coverage for all age groups. Secondly, we found an excellent positive predictive value (90%–100%) in the inpatient setting, validating the ICD-9-CM code 691 for AD (23). We also qualified the definitions of asthma and AD in our study to include a minimum of three outpatient records or one inpatient record. Our database size ensured similar distributions in each group due to well-balanced matching, which reduced the study's heterogeneity and selection bias. To further reduce bias, we performed a sensitivity analysis for unmeasured confounders. Importantly, this is the first population-based cohort study to show that influenza vaccination could reduce the incidence of asthma in AD patients.

There are some limitations to this study, particularly since there is no consensus on asthma's diagnosis, especially in children. The ICD-9-CM code 493–based algorithm for ascertaining asthma had a sensitivity of 82% and specificity of 98% (24), and a positive predictive value of 75.0% when compared to the criteria-based medical record review (25). The ICD-9-CM code 493 is widely accepted for etiologic research in asthma, but may underestimate its prevalence. Further efforts are needed to check the consistency of diagnosis by medical chart review. In order to reduce confounding by indication, an active comparator (i.e., other vaccines) is needed to serve as a control group. Finally, the clinical relevance of this study must be further validated by larger-scale prospective randomized trials.

CONCLUSION

This long-term nationwide cohort study revealed that influenza vaccination was associated with lower incidental asthma risk in people with AD after adjusting for age, gender, and comorbidities. Nevertheless, more comprehensive studies are needed to confirm our findings.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors contributed to the article and approved the submitted version. Conception and design: KHL and JC-CW. Acquisition of data: YHW. Analysis and interpretation of data: KHL, YHW, and JC-CW. Writing (original draft preparation): KHL. Writing (review and editing): P-YL, C-FT, and JC-CW.

FUNDING

This study was supported by Chung Shan Medical University Hospital (CSH-2019-C-004). The funders had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; or the preparation, review, or approval of the manuscript.

REFERENCES

1. Eichenfield LF, Hanifin JM, Beck LA, Lemanske RF, Sampson HA, Weiss ST, et al. Atopic Dermatitis and Asthma: Parallels in the Evolution of Treatment. *Pediatrics* (2003) 111(3):608–16. doi: 10.1542/peds.111.3.608
2. Naldi L, Parazzini F, Gallus S. Prevalence of Atopic Dermatitis in Italian Schoolchildren: Factors Affecting Its Variation. *Acta Derm Venereol* (2009) 89 (2):122–5. doi: 10.2340/00015555-0591
3. Belgrave DC, Granell R, Turner SW, Curtin JA, Buchan IE, Le Souëf PN, et al. Lung Function Trajectories From Pre-School Age to Adulthood and Their Associations With Early Life Factors: A Retrospective Analysis of Three Population-Based Birth Cohort Studies. *Lancet Respir Med* (2018) 6(7):526–34. doi: 10.1016/S2213-2600(18)30099-7
4. Cho Y, Kim TB, Lee TH, Moon KA, Lee J, Kim YK, et al. Chlamydia Pneumoniae Infection Enhances Cellular Proliferation and Reduces Steroid Responsiveness of Human Peripheral Blood Mononuclear Cells *via* a Tumor Necrosis Factor-α-Dependent Pathway. *Clin Exp Allergy* (2005) 35(12):1625–31. doi: 10.1111/j.1365-2222.2005.02391.x
5. Wood LG, Simpson JL, Hansbro PM, Gibson PG. Potentially Pathogenic Bacteria Cultured From the Sputum of Stable Asthmatics are Associated With Increased 8-Isoprostane and Airway Neutrophilia. *Free Radical Res* (2010) 44 (2):146–54. doi: 10.3109/10715760903362576
6. Hansbro NG, Horvat JC, Wark PA, Hansbro PM. Understanding the Mechanisms of Viral Induced Asthma: New Therapeutic Directions. *Pharmacol Ther* (2008) 117(3):313–53. doi: 10.1016/j.pharmthera. 2007.11.002
7. Busse WW, Lemanske RFJr., Gern JE. Role of Viral Respiratory Infections in Asthma and Asthma Exacerbations. *Lancet* (2010) 376(9743):826–34. doi: 10.1016/S0140-6736(10)61380-3
8. Fuchs O, von Mutius E. Prenatal and Childhood Infections: Implications for the Development and Treatment of Childhood Asthma. *Lancet Respir Med* (2013) 1(9):743–54. doi: 10.1016/S2213-2600(13)70145-0
9. Dawood FS, Kamimoto L, D'Mello TA, Reingold A, Gershman K, Meek J, et al. Children With Asthma Hospitalized With Seasonal or Pandemic Influenza, 2003–2009. *Pediatrics* (2011) 128(1):e27–32. doi: 10.1542/peds.2010-3343
10. Merckx J, Ducharme FM, Martineau C, Zemek R, Gravel J, Chalut D, et al. Respiratory Viruses and Treatment Failure in Children With Asthma Exacerbation. *Pediatrics* (2018) 142(1):e20174105. doi: 10.1542/peds.2017-4105
11. Vasileiou E, Sheikh A, Butler C, El Ferkh K, Von Wissmann B, McMenamin J, et al. Effectiveness of Influenza Vaccines in Asthma: A Systematic Review and Meta-Analysis. *Clin Infect Dis* (2017) 65(8):1388–95. doi: 10.1093/cid/cix524
12. Kim BE, Leung DY, Boguniewicz M, Howell MD. Loricrin and Involucrin Expression Is Down-Regulated by Th2 Cytokines Through STAT-6. *Clin Immunol* (2008) 126(3):332–7. doi: 10.1016/j.clim.2007.11.006
13. Marsland B, Scanga C, Kopf M, Le Gros G. Allergic Airway Inflammation is Exacerbated During Acute Influenza Infection and Correlates With Increased Allergen Presentation and Recruitment of Allergen-Specific T-Helper Type 2 Cells. *Clin Exp Allergy* (2004) 34(8):1299–306. doi: 10.1111/j.1365-2222.2004.02021.x
14. Scherle P, Palladino G, Gerhard W. Mice can Recover From Pulmonary Influenza Virus Infection in the Absence of Class I-Restricted Cytotoxic T Cells. *J Immunol* (1992) 148(1):212–7.
15. Graham MB, Braciale VL, Braciale TJ. Influenza Virus-Specific CD4+ T Helper Type 2 T Lymphocytes do Not Promote Recovery From Experimental Virus Infection. *J Exp Med* (1994) 180(4):1273–82. doi: 10.1084/jem.180.4.1273
16. Maloy KJ, Burkhart C, Junt TM, Odermatt B, Oxenius A, Piali L, et al. CD4+ T Cell Subsets During Virus Infection: Protective Capacity Depends on Effector Cytokine Secretion and on Migratory Capability. *J Exp Med* (2000) 191 (12):2159–70. doi: 10.1084/jem.191.12.2159
17. Skevaki C, Hudemann C, Matrosovich M, Möbs C, Paul S, Wachtendorf A, et al. Influenza-Derived Peptides Cross-React With Allergens and Provide Asthma Protection. *J Allergy Clin Immunol* (2018) 142(3):804–14. doi: 10.1016/j.jaci.2017.07.056
18. McElhaney JE. Influenza Vaccine Responses in Older Adults. *Ageing Res Rev* (2011) 10(3):379–88. doi: 10.1016/j.arr.2010.10.008
19. Barnetson RSC, Rogers M. Childhood Atopic Eczema. *Bmj* (2002) 324 (7350):1376–9. doi: 10.1136/bmj.324.7350.1376
20. Mirabelli MC, Beavers SF, Chatterjee AB, Moorman JE. Age at Asthma Onset and Subsequent Asthma Outcomes Among Adults With Active Asthma. *Respir Med* (2013) 107(12):1829–36. doi: 10.1016/j.rmed.2013.09.022
21. Asher MI, Montefort S, Björkstén B, Lai CK, Strachan DP, Weiland SK, et al. Worldwide Time Trends in the Prevalence of Symptoms of Asthma, Allergic

Rhinoconjunctivitis, and Eczema in Childhood: ISAAC Phases One and Three Repeat Multicountry Cross-Sectional Surveys. *Lancet* (2006) 368(9537):733–43. doi: 10.1016/S0140-6736(06)69283-0

22. Cazzola M, Calzetta L, Bettoncelli G, Novelli L, Cricelli C, Rogliani P. Asthma and Comorbid Medical Illness. *Eur Respir J* (2011) 38(1):42–9. doi: 10.1183/09031936.00140310
23. Hsu DY, Dalal P, Sable KA, Voruganti N, Nardone B, West D, et al. Validation of International Classification of Disease Ninth Revision Codes for Atopic Dermatitis. *Allergy* (2017) 72(7):1091–5. doi: 10.1111/all.13113
24. Seol HY, Wi C-I, Ryu E, King KS, Divekar RD, Juhn YJ. A Diagnostic Codes-Based Algorithm Improves Accuracy for Identification of Childhood Asthma in Archival Data Sets. *J Asthma* (2020) 1–10. doi: 10.1080/02770903.2020.1759624
25. Juhn Y, Kung A, Voigt R, Johnson S. Characterisation of Children's Asthma Status by ICD-9 Code and Criteria-Based Medical Record Review. *Primary Care Respir J* (2011) 20(1):79–83. doi: 10.4104/pcrj.2010.00076

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Immunology

ORIGINAL RESEARCH
published: 18 November 2021
doi: 10.3389/fimmu.2021.729681

Proteomic Approaches to Defining Remission and the Risk of Relapse in Rheumatoid Arthritis

Liam J. O'Neil[1,2*], *Pingzhao Hu*[3,4], *Qian Liu*[3,4], *Md. Mohaiminul Islam*[3,4], *Victor Spicer*[2], *Juergen Rech*[5], *Axel Hueber*[5], *Vidyanand Anaparti*[2], *Irene Smolik*[1], *Hani S. El-Gabalawy*[1,2], *Georg Schett*[5†] and *John A. Wilkins*[1,2†]

[1] Section of Rheumatology, Department of Internal Medicine, University of Manitoba, Winnipeg, MB, Canada, [2] Manitoba Centre for Proteomics and Systems Biology, University of Manitoba and Health Sciences Centre, Winnipeg, MB, Canada, [3] Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, MB, Canada, [4] Department of Computer Science, University of Manitoba, Winnipeg, MB, Canada, [5] Department of Medicine, Friedrich-Alexander University Erlangen-Nuernberg and Universitaetsklinikum Erlangen, Erlangen, Germany

OPEN ACCESS

Edited by:
Xu-jie Zhou,
Peking University First Hospital, China

Reviewed by:
Loredana Frasca,
National Institute of Health (ISS), Italy
Gillian Inara Rice,
The University of Manchester,
United Kingdom
Feiyang Ma,
University of California, Los Angeles,
United States

****Correspondence:***
Liam J. O'Neil
liam.oneil@umanitoba.ca

†These authors have contributed equally to this work

Specialty section:
This article was submitted to
Autoimmune and Autoinflammatory Disorders,
a section of the journal
Frontiers in Immunology

Received: *23 June 2021*
Accepted: *20 October 2021*
Published: *18 November 2021*

Citation:
O'Neil LJ, Hu P, Liu Q, Islam MM, Spicer V, Rech J, Hueber A, Anaparti V, Smolik I, El-Gabalawy HS, Schett G and Wilkins JA (2021) Proteomic Approaches to Defining Remission and the Risk of Relapse in Rheumatoid Arthritis. Front. Immunol. 12:729681. doi: 10.3389/fimmu.2021.729681

Objectives: Patients with Rheumatoid Arthritis (RA) are increasingly achieving stable disease remission, yet the mechanisms that govern ongoing clinical disease and subsequent risk of future flare are not well understood. We sought to identify serum proteomic alterations that dictate clinically important features of stable RA, and couple broad-based proteomics with machine learning to predict future flare.

Methods: We studied baseline serum samples from a cohort of stable RA patients (RETRO, n = 130) in clinical remission (DAS28<2.6) and quantified 1307 serum proteins using the SOMAscan platform. Unsupervised hierarchical clustering and supervised classification were applied to identify proteomic-driven clusters and model biomarkers that were associated with future disease flare after 12 months of follow-up and RA medication withdrawal. Network analysis was used to define pathways that were enriched in proteomic datasets.

Results: We defined 4 proteomic clusters, with one cluster (Cluster 4) displaying a lower mean DAS28 score ($p = 0.03$), with DAS28 associating with humoral immune responses and complement activation. Clustering did not clearly predict future risk of flare, however an XGboost machine learning algorithm classified patients who relapsed with an AUC (area under the receiver operating characteristic curve) of 0.80 using only baseline serum proteomics.

Conclusions: The serum proteome provides a rich dataset to understand stable RA and its clinical heterogeneity. Combining proteomics and machine learning may enable prediction of future RA disease flare in patients with RA who aim to withdrawal therapy.

Keywords: rheumatoid arthritis, disease activity, outcomes research, treatment, proteomics

HIGHLIGHTS

- Serum proteomics defines clinically relevant clusters within a cohort of stable RA patients
- Machine learning and proteomics may identify individuals at highest risk for future disease flare
- Despite meeting criteria for remission, clinically detectable disease is associated with a serum proteomic signature in stable RA

INTRODUCTION

Rheumatoid Arthritis (RA) is a systemic autoimmune disease that is characterized by inflammation of synovial joints (1). Modern RA therapy is initiated early and escalated aggressively using a treat-to-target approach to try an obtain disease remission (2). The development of both targeted treatments and combination regimens continues to improve expected outcomes for patients. Encouragingly, clinical remission, defined by multiple measures of disease activity (3), has become a realistic expectation for most patients with RA. Recent registry data of RA cohorts consistently show that DAS28 (Disease Activity Score) remission is achieved in about 50% of patients (4), a number that may be increasing over time (5).

Patients with RA who are able to achieve disease remission using standard therapy are not well studied, given their lack of disease activity and need for treatment changes. The main issue facing these patients is whether or not to remain on their treatment, or risk withdrawal and the potential for disease flare. There are many prospective studies that have demonstrated successful Disease Modifying Anti-Rheumatic Drugs (DMARD) withdrawal in patients in clinical remission (6–9) but the determinants of maintaining remission status after medication withdrawal are poorly defined (10). Unfortunately, given the limited understanding of the pathological mechanisms that drive subclinical disease, clinicians are left to guess which of their patients might sustain remission using less aggressive therapy.

Technological advances in high-throughput proteomics have allowed for an improved understanding of disease processes and biomarker discovery (11). Although mass spectrometry tends to dominate this evolving field, broad-based targeted proteomics has its own advantages, including simplified sample preparation and user-friendly output data (12). Our group has previously defined protein sets that are associated with future disease flare from pre-clinical RA by coupling machine learning with proteomic approaches (13). Indeed, leveraging omics approaches to resolve heterogeneity in common diseases remains a distinct challenge in clinical medicine (14), though this has not been systematically undertaken in a stable RA cohort.

The RETRO (15) (Reducing therapy in rheumatoid arthritis patients in ongoing remission) study is a prospective randomized trial which enrolled patients who had achieved disease remission with conventional RA therapy. One of the aims of this study is to define disease recurrence in patients with RA when either continuing or reducing their medications. It was previously shown in this trial that positive anti-citrullinated antibody (ACPA), and other biomarkers (16, 17) are associated with an increased likelihood of disease relapse. In spite of these studies, there is little understanding of the underlying biological mechanisms that are active in stable RA. If differences within RA patients in remission can be more clearly defined, there may be an enhanced understanding of the spectrum of RA pathogenesis, along with improved personalized clinical approaches surrounding the withdrawal of therapy. We hypothesized that high-throughput proteomics (18) would help identify underlying biological heterogeneity that might provide insights into mechanisms underpinning future disease flare. Our aim was to explore how the serum proteome shapes the underlying clinical experiences of stable RA patients.

METHODS

Patients and Inclusion Criteria

RETRO is a multicentre, randomized, open, prospective, controlled parallel-group study. Details of the study are described in the original publication (15). The objective of the study is to evaluate tapering or discontinuation of DMARDs in patients with RA. All enrolled patients fulfilled the 2010 American College of Rheumatology (ACR) criteria for RA (19). Patients had to have sustained clinical remission defined by the Disease Activity score (DAS28 < 2.6) criteria for at least 6 months (20). Ethics committee of the Friedrich-Alexander-University of Erlangen-Nuremberg approval was granted.

Treatment and Follow-up

Patients were randomized to one of three arms: Arm 1 continued with existing DMARD regimen at full dose for 12 months, arm 2 reduced the dose of all DMARDs by 50%, while arm 3 reduced the dose of all DMARDs by 50% in the first 6 months, then discontinued all medications. Relapse of disease was defined as a DAS28-ESR score greater than 2.6. Participants were assessed for clinical disease activity every 3 months until month 12.

Assessment of Demographic and Disease-Specific Parameters

Age and sex were recorded in all patients. Disease duration, tender joint count (68), swollen joint count (66), patient visual analogue scale (VAS) for pain and patient global were assessed and recorded. C-reactive protein (CRP), ESR, Rheumatoid Factor (RF), ACPA, DAS28-ESR and Health Assessment Questionnaire (HAQ-DI) were recorded.

SOMAscan

SOMAscan is a proteomics assay that measures 1307 proteins using an aptamer library. This high-throughput proteomics assay has been used in recent publications to study the aging proteome (21, 22) along with other human diseases (23). 130 baseline serum samples were available from the RETRO study. Briefly, a library of aptamers were incubated with serum, and those that bind are isolated and hybridized to DNA microarray for detection. The identity and relative concentration of the detected proteins are revealed by localization and fluorescence intensity. Protein quantification is reported as relative fluorescence units (RFU), an arbitrary value. In general, agreement between aptamer and antibody-based assays is high (24). Further details regarding the SOMAscan assay are available (18).

Statistical Analysis

Descriptive results (**Table 1**) are stated in means and standard deviation. SOMAmer protein expression RFU values for the study patients were transformed into a log2 scale for differential analysis (**Supplemental File 2**). Batch effect was removed in our SOMAmer data using internal controls within each plate to adjust proteomic intensity as per standard SOMAscan protocols. Batch effect was assessed between plates and determined to require no further correction. Data was loaded and analyzed in the R (v3.5.3) environment unless otherwise stated. Missing clinical data was imputed using multiple imputation by chained equations (MICE) (25). Differential analysis between groups was undertaken using linear modeling with the package *LIMMA (*26). GO pathways analysis was performed using clusterprofiler (27). Graphs were generated using the *ggplot2* package. Correlation analyses were performed for select proteins using Pearson correlation. Multi-dimensional scaling (MDS) was used for dimension reduction on all SOMAscan proteins.

The 200 most variable proteins measured by coefficient of variation were used to determine optimal number of clusters ranging from k = 2 to 10 and identify sample clusters using the R package *Consensusclusterplus*. We used 80% protein resampling and 80% patient resampling and selected Pearson as our distance function. Multinomial logistic regression implemented in R package *glmnet* was used to identify clinical variables that are independently associated with cluster assignment. Sliding window analysis of DAS28 scores and protein expression was performed using a previously published algorithm, DE-SWAN (28). Briefly, this algorithm analyzed serum protein expression across quintiles of DAS28 scores using linear modeling, while adjusting for baseline demographic factors, in this case age and sex. A protein expression score was developed on proteins that correlated with DAS28 which were identified by DE-SWAN. We filtered the proteomic data on the 34 score members, scaled the data by the mean and standard deviation, and multiplied by 1 (positively associated with DAS28) or -1 (negatively associated with DAS28) for each protein. The final score was the mean expression of all 34 proteins for each patient. We randomly generated 5000 data sets with 34 randomly selected proteins in each set to evaluate the significance of the association score.

Machine Learning Classification Algorithm

We applied two supervised machine learning (ML) techniques to develop algorithms to classify flare or remission based on serum proteomics. The first approach we used is XGBoost (Extreme Gradient Boosting), which employs a regularization term to overcome the overfitting (29). The second approach is the LASSO model (30), which was used as a baseline to compare its performance with that of XGBoost. Data was loaded into Python, and samples were randomly split into a training (n = 104, 80% of the samples) and test (n = 26) set. The training set was used to train and tune the parameters in the two models and the test set was used evaluate the models' performance, which were measured by the area under of the curve (AUC) of receiver operating characteristic (ROC), accuracy, sensitivity and specificity. To increase the interpretability of the XGBoost model to predict the flare status of a given sample, we used SHAP values (Shapley Additive Explanation) (31). A higher SHAP value of a given feature in the model represents its strong influence on the model output. The final model parameters we used in the XGBoost are as follows: learning_rate = 0.01, max_depth = 3, subsample = 0.6, colsample_bytree = 0.7, n_estimators = 100, gamma = 0.0, reg_alpha = 0.5, the parameters used in the LASSO model is as follow: cost=1.17 and max_iterations = 5000. We used 5-fold cross-validation to get the optimal hyperparameters.

TABLE 1 | Baseline characteristics of the patients, split by proteomic cluster.

Characteristics	Total (n = 130)	Cluster 1 (n = 34)	Cluster 2 (n = 12)	Cluster 3 (n = 46)	Cluster 4 (n = 38)
Age	55.2 (13.1)	52.7 (14.6)	54.1 (13.7)	54.7 (11.7)	58.6 (13.1)
Females, %	56.2%	67.6%	66.7%	56.5%	42.1%
Disease Duration (years)	6.8 (7.0)	7.9 (6.5)	8.6 (9.3)	6.9 (6.3)	4.9 (7.3)
DAS-28 (ESR)	1.7 (0.68)	1.93 (0.60)	1.73 (0.74)	1.71 (0.65)	1.51 (0.71)
ACR/EULAR remission, %	76.6%	67.7%	66.6%	88.9%	72.9%
HAQ, units	0.12 (0.32)	0.11 (0.17)	0.08 (0.12)	0.16 (0.46)	0.09 (0.26)
Positive RF, %	56.2%	73.5%	50.0%	52.2%	47.3%
Positive ACPA, %	57.7%	67.6%	66.7%	55.6%	50.0%
*Biological DMARD use, % (N)	40.0%	38.2%	25.0%	47.8%	36.8%
Flare, %	37.7%	38.2%	16.7%	43.5%	36.8%

ACPA, anticitrullinated protein antibody; ACR, American College of Rheumatology; CRP, C-Reactive protein; DAS-28, disease activity score-28 (based on ESR); DMARDs, disease modifying antirheumatic drugs; ESR, erythrocyte sedimentation rate; EULAR, European League Against Rheumatism; HAQ, Health Assessment Questionnaire; RF, Rheumatoid Factor; VAS, Visual analogue scale.

**Tumor necrosis factor inhibitors and tocilizumab.*

Study Cohort

Baseline characteristics for 130 patients enrolled in the RETRO study are found in **Table 1**. Overall, the group had maintained clinical remission for 16.6 (16.2) months and mean disease duration of over 6 years. 57.7% of the patients were ACPA positive, while 40.0% required biologics to achieve remission. 76.6% of patients had achieved the most stringent definition of remission [ACR/EULAR remission (32)]. After 12 months of follow-up, 62.3% of the overall population remained in clinical remission (50% in those undergoing withdrawal).

RESULTS

Hierarchical Clustering on Serum Proteins Identifies Heterogeneity Amongst Stable RA Patients

Given the paucity of data aimed at understanding subclinical disease activity in RA patients who achieve remission, we sought to explore underlying heterogeneity using serum proteomics in this established cohort. We quantified over 1300 serum proteins from 130 RETRO patients at their baseline visit, all of whom were in stable clinical remission (DAS28 < 2.6). We hypothesized that despite the clinical similarities amongst individuals within this cohort, proteomic differences may provide important insights by identifying sub-clusters of patients. We applied consensus clustering to assign individuals to one of the 4 clusters (**Figure S1**) and clustered scaled protein expression by hierarchical clustering, which can be seen in **Figure 1**. MDS analysis revealed separation of the hierarchical clusters (**Figure S2**).

Baseline characteristics split by cluster are listed in **Table 1**. We found no differences in sex, age, biologic use, or serological status across our 4 proteomic clusters. Cluster 4 had significantly lower DAS28 scores compared to the remaining clusters. With respects to future flare, Cluster 2 trended toward lower rates relative to the remaining clusters, however this did not reach statistical significance (16.7% vs 39.8%, p = 0.21). To assess this association by multinomial regression, we assigned Cluster 2 as the reference cluster and found that that Cluster 3 had higher odds of flare (OR 5.6, 0.97 to 33.06, p = 0.05), relative to Cluster 2 with similar trends observed for Cluster 1 and Cluster 4 (**Table S1**). Indeed, no clear distinction between individuals who developed future flare was observed in the MDS plot (**Figure S2**). Overall, these results suggest that global proteomic clusters within a clinically homogenous cohort can be identified but are associated with current clinical status rather than future outcomes.

Machine Learning Classifies Future Flare Using Baseline Serum Proteomics

We next aimed to use the serum proteome to identify biomarkers associated with future disease flare in stable RA, given that clustering did not clearly associate with risk of flare. We identified DEP's between these groups (**Table S2**) and observed

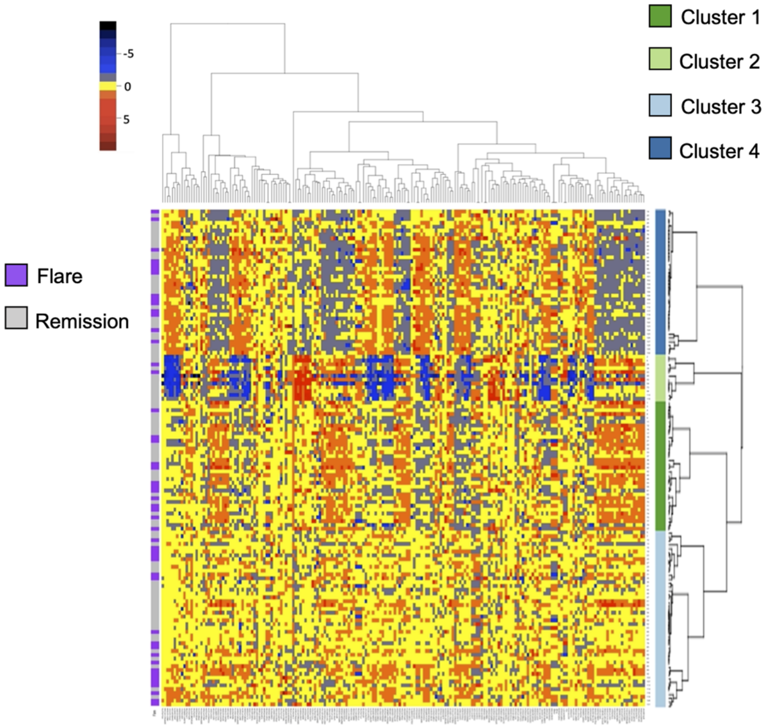


FIGURE 1 | Heatmap and hierarchical clustering of 200 serum proteins in stable RA. Protein expression is scaled and colored by relative expression. Each column is a protein and each row is a patient. Clustering is shown in both dimensions. Flare or remission is annotated in purple or grey to the left of the graph.

upregulation in Ectodysplasin A receptor (EDAR, FC = 1.20) and Serine peptidase inhibitor (SPINT2, FC 1.1), and downregulation of Fractalkine (CX3CL1, FC 0.95) and Ephrin type-B receptor 2 (EPHB2, FC 0.95) in individuals who eventually went on to flare (**Figure 2A**). However, after adjustment for multiple comparisons, none of the differentially expressed proteins reached statistical significance. This suggests that although subtle differences exist in the serum proteome between individuals who experience future flare, it's unlikely that singular biomarkers accurately predict this outcome in this population.

To test this hypothesis, we explored the use of two machine learning algorithms, LASSO and XGBoost, to build predictive models that classify future flare using baseline serum proteomics. We generated two models, both of which were validated on a test

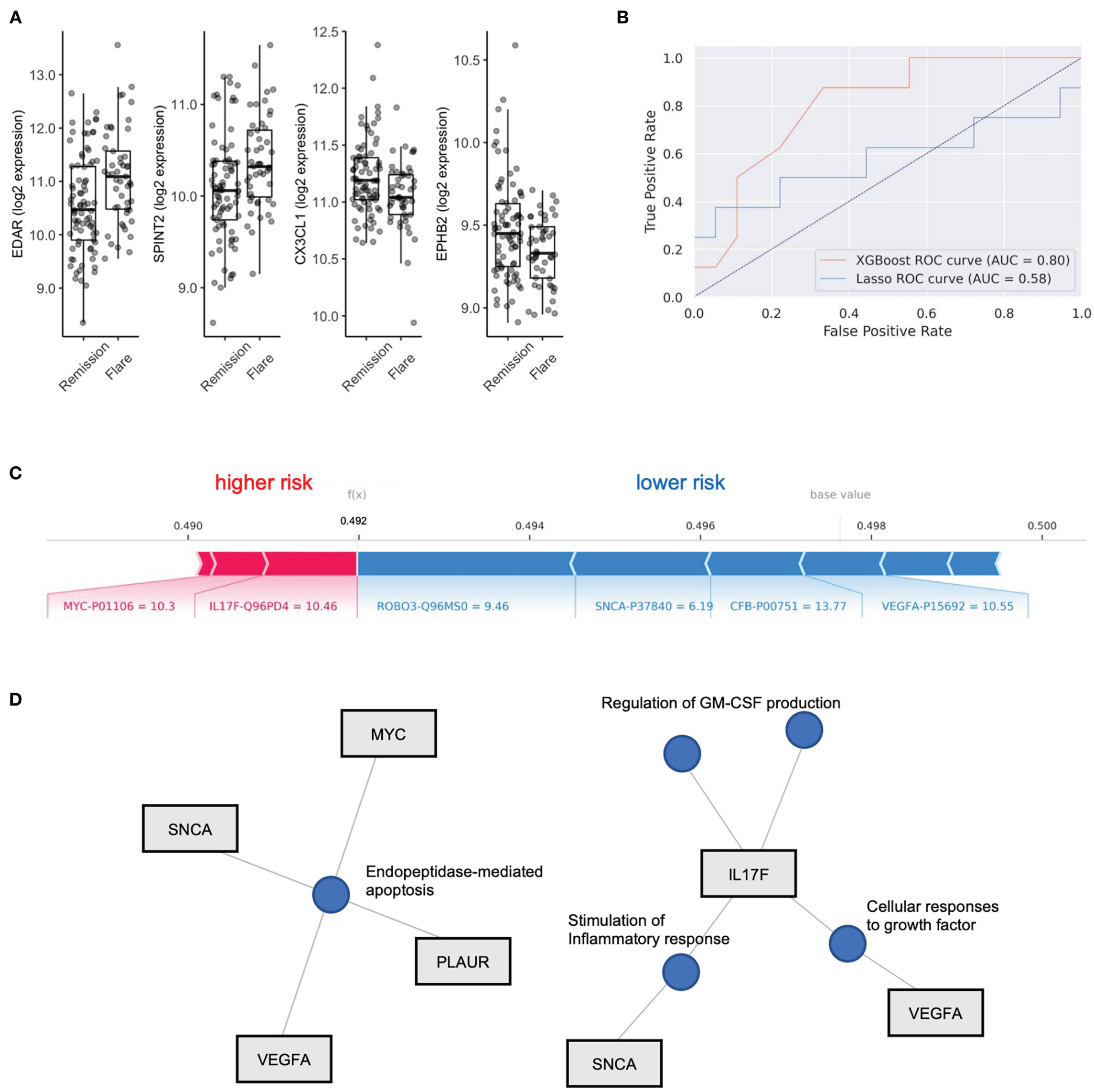


FIGURE 2 | XGboost machine learning to identify flare or remission in stable RA patients. **(A)** Box plots of EDAR, SPINT2, CX3CL1 and EPHB2 split by individuals who remained in Remission or Flare. **(B)** Receiver operator curves (ROC) of 2 machine learning models, XGboost and LASSO, trained on serum proteome to classify flare or remission. AUC is representative of test set cohort parameters. **(C)** Bar plot of model features that impact risk of flare or remission in the XGboost model with log2 expression and Uniprot ID annotated for each protein member. Feature importance is represented by relative size of bar. Values for different proteins represent their original values in the dataset for that particular sample. The base value means the average of the prediction scores, and 0.5 is the cutoff threshold to select a Flare status. **(D)** Gene concept plots derived from XGboost protein features. Each node represents a GO pathway with proteins connected by edges. EDAR, Ectodysplasin A receptor; SPINT2, Serine peptidase inhibitor; CX3CL1, Fractalkine; EPHB2, Ephrin type-B receptor 2. SNCA, Synuclein alpha; PLAUR, Plasminogen Activator; VEGFA, vascular endothelial growth factor A; MYC, Myc proto-oncogene protein; IL17F, Interleukin 17F; ROBO3, Roundabout homolog 3; CFB, complement factor B.

cohort (20% of total cohort). The LASSO model achieved 69.2% accuracy, with an AUC of 0.58, along with sensitivity of 0.5 and specificity of 0.78 based on the test cohort (Features in **Table S3**). XGboost delivered a model with higher specificity (0.78) than sensitivity (0.63) and an overall accuracy of 73.1% with an AUC 0.80. Therefore, we found that the XGboost model outperformed the LASSO model by the metric area under the curve (**Figure 2B**, AUC, 0.80 *vs* 0.58), accuracy (73.1% *vs* 69.2%) and sensitivity (0.63 *vs.* 0.5).

To interpret this XGBoost model, we explored the impact of essential features in terms of SHAP values on the classifier's output for a single prediction which are shown in **Figure 2C**. We identified Interleukin 17F (IL17F) and Myc proto-oncogene protein (MYC) expression as indicators of future flare, while Roundabout homolog 3 (ROBO3), Synuclein alpha (SNCA), complement factor B (CFB) and vascular endothelial growth factor A (VEGF-A) expression were indicators of sustained remission (**Figure 2C** and **Figure S3**). Given the small number of proteins that derived our boosted model, we next explored whether there were any functional links between these proteins. We developed gene concept plots to identify potential protein interactions and found that SNCA, MYC, VEGFA and PLAUR were connected by a single pathway, *endopeptidase mediated apoptosis*. We then analyzed IL17F restricted networks, given its conflicting role in RA (33–36), and that its expression was associated with future flare in our model. We found that IL17F interacted with VEGFA though *growth factor* function, and with SNCA through common effects on the *inflammatory response* (**Figure 2D**). IL17F independently regulated *GM-CSF production*, a key driver of RA disease activity through the recruitment of neutrophils (37). Overall, this network analysis suggests that cellular apoptosis and GM-CSF production may be associated with future disease flares in RA patients who are otherwise stable.

Disease Activity in Stable RA Is Reflected in the Serum Proteome

In our hierarchical clustering, we observed a lower mean DAS28 score in Cluster 4 compared to the remaining 3 clusters (**Figure 3A**). RA patients who achieve DAS28 defined remission often have residual disease activity, however, since this population is not typically the focus of translational studies, little is known regarding biomarkers that are reflective of ongoing disease activity. Indeed, we found that several protein members correlated with DAS28 score (**Figure S4**), including Integrin alpha 2B (ITGA2B), Bactericidal permeability-increasing protein (BPI) and chemokine ligand 2 (CXCL2, Pearson R, all p value < 0.01). Further, complement proteins (C3, C4A, C1S) were all negatively associated with DAS28 score, suggesting activation and consumption of complement proteins (**Figure 3B**) were indicators of disease. There was no indication that these parameters varied based on ACPA status (**Figure S5**).

To further explore the relationship between disease activity and serum biomarkers, we developed a sliding window model (SWAN) which examined protein variability across DAS28 quintiles, after controlling for *Age* and *Sex*. Across DAS28 a total of 34 proteins varied significantly with disease activity (**Figure 4A**). We used these 34 protein members to annotate a meta-protein expression score (the mean expression profiles of the 34 proteins), which correlated with DAS28 (R = 0.45, p < 0.001, **Figure S6**). To test the robustness of this finding, we sampled 5000 random sets with 34 proteins in each set and correlated their mean expression with DAS28 scores. We found a range of -0.37 – 0.27, associated with a low probability (0) that the correlation of 0.45 would occur by chance (**Figure S7**). This protein disease activity score was significantly lower in Cluster 4 compared to the remaining 3 clusters (**Figure 4B**), concordant with their lower DAS28 scores. Gene concept plots revealed that

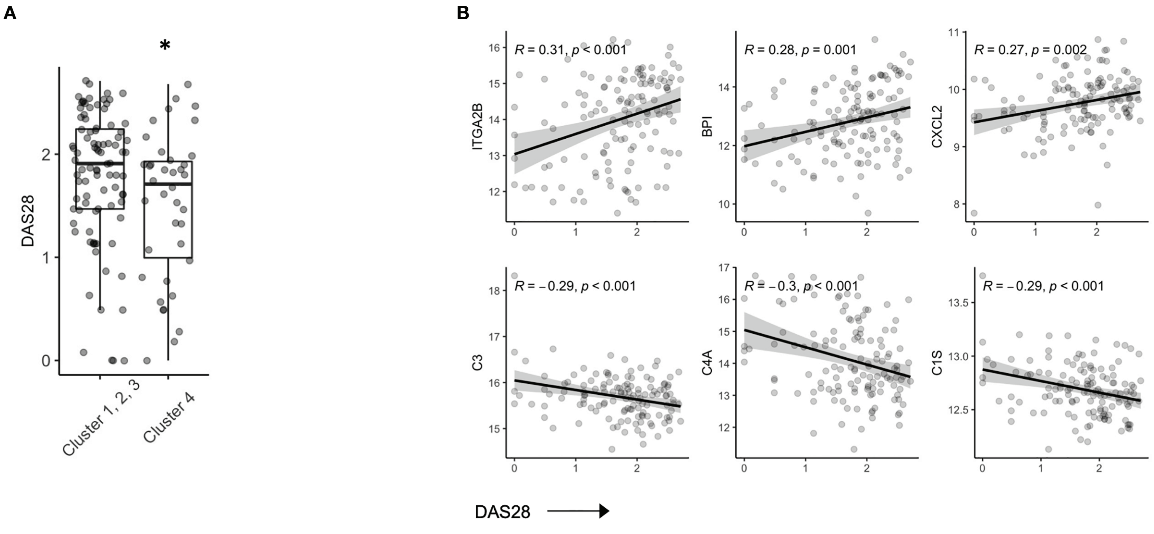


FIGURE 3 | Serum proteins are associated with DAS28 in stable RA. **(A)** Box plots of DAS28 disease scores in patients, split by cluster assignment (p = 0.03). **(B)** Correlation plots of DAS28 and serum protein expression of ITGA2B, BPI, CXCL2, C3, C4A, C1S. *p < 0.05.

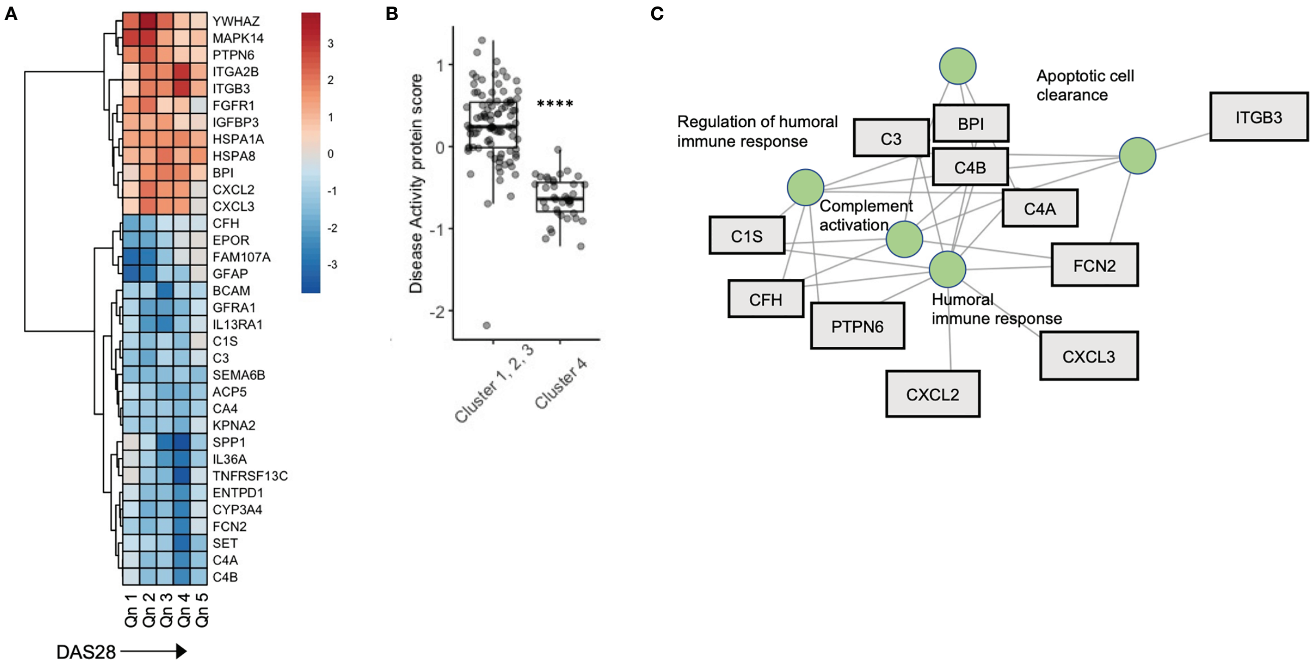


FIGURE 4 | A serum proteomic signature is associated with DAS28 in stable RA. **(A)** Sliding window analysis of disease activity (Quintiles x-axis) and 34 proteins that vary with DAS score after controlling for *Age* and *Sex*. Heatmap is colored on coefficient relationship to DAS score **(B)** Box plots of disease activity protein score split by cluster assignment. **(C)** Gene concept plots derived from disease activity protein score. Each node represents a GO pathway with proteins connected by edges. ITGA2B, Integrin alpha 2B; BPI, Bactericidal permeability-increasing protein; CXCL2, Chemokine ligand 2; C3, Complement C3; C4A, Complement 4A; C1S, Complement 1S. ****p < 0.00001.

these 34 proteins interacted through nodes that included *humoral immunity, apoptosis*, and *complement activation* (**Figure 4C**). Overall, these data suggest that stable RA disease activity is marked by complement consumption and humoral immune responses, which is reflected in a serum protein signature that is detectable in patients with stable RA.

DISCUSSION

With the advent of multiple targeted therapies, alone and in combination, most RA patients should reasonably expect to achieve low disease activity or clinical remission status. To date, few studies have sought to understand the heterogeneity of the biological pathways underpinning clinical remission in this rapidly expanding population of RA patients. When, and in whom to attempt withdrawal of therapy has become a compelling clinical question. On the one hand, there is justifiable concern regarding reactivation of systemic and articular inflammation. On the other hand, ongoing use of DMARDs and/or biologics is associated with increased risk of infectious complications (38), malignancy (39) and cost (40). Strategies for successful taper however remain ill-defined and, importantly, lack precision (6, 41, 42). The RETRO clinical trial has previously generated predictive models that were based on clinical parameters and serum studies (16, 17). ACPA seropositivity appears to be an important indicator for increased risk of future relapse (17), while clinical parameters have modest predictive value even when combined with advanced machine learning techniques (43). It remains unclear if these indicators are clinically applicable and generalizable to a wide range of RA patient populations. The focus of this study was to use a broad-based serum proteomic approach to better understand the underlying heterogeneity amongst RA patients who are in sustained clinical remission, prior to their participation in a clinical trial of therapy withdrawal. Our results identify proteomic signatures reflecting biological mechanisms that are associated with ongoing disease stability off therapy, or alternatively, the risk of future disease relapse.

Our XGboost model suggested that individual circulating serum biomarkers are unlikely, on their own, to be predictive of future stability or relapse after therapy withdrawal. In spite of this, combinations of proteomic biomarkers identified by the machine learning achieved relatively high AUC and accuracy in predicting outcomes. Indeed, this is a testament to the power of rapidly evolving machine learning algorithms that are being developed for many clinical problems (44). Network analysis of proteins derived from machine learning suggested that inflammatory forms of cellular death was an indicator for risk future disease flare. Indeed, apoptosis is escaped by pathogenic fibroblast-like synoviocytes and likely contributes to their aggressive and hyperplastic phenotype in RA (45) and this may point to systemic FLS as a potential source of these proteins (46). Hierarchical clustering identified proteomic clusters which were defined, in part, by clinical characteristics. Cluster 4 represented a patient group with lower DAS28 scores amongst the remaining patient cohort. Our disease activity signature, found to be lower in Cluster 4, suggested that elements of humoral immunity and complement activation might facilitate disease activity in otherwise stable RA patients. This suggests that activating pathways differentially regulate disease activity in this subset of RA patients, as many of the well-known disease activity markers in RA suggest that innate immune responses associate with DAS28 scores (47–49).

The analyses we undertook utilized the SOMAscan aptamer-based technology to interrogate 1307 distinct serum proteins. Although this provided us with a robust array of biomarkers, it is well recognized that these represent only a fraction of the human proteome, and that larger arrays that span a larger proportion of the circulating proteome may help generate even more accurate predictive algorithms. Moreover, there remains an incomplete understanding of how aptamer-based detection of each individual analyte correlates with other detection methodologies such as those that are antibody based (24). Due to our modest sample size, we observed strongly statistically significant association although R values related to DAS28 scores were all below 0.3. However, we expect the R values may increase using a larger sample size while the significant association will be still held. Finally, SOMA proteins may bias over-representation analysis based on the selected proteins which are included in the set. Notably, network analysis was used in this study to connect proteins of interest through biological nodes. These results would not be impacted by inherent bias in the SOMA protein set.

In conclusion, we applied an unsupervised, high-throughput proteomics assay to delineate biomarkers and pathways that reflect the biological heterogeneity present in RA patients who are collectively deemed to be in stable clinical remission. Based on this, we used supervised machine learning to develop robust models that predicted ongoing disease stability after therapy withdrawal as opposed to future disease flare. Although it is premature to try and define the potential clinical utility of these models, they do provide an important impetus for further studies that aim to further define a biological definition of remission in RA patients that can ultimately guide clinical decision making.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Friedrich-Alexander-University of Erlangen-Nuremberg. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LO'N: data analysis, manuscript construction. PH: data analysis. QL: data analysis. MI: data analysis. VS: data analysis. JR: data collection and clinical trial. AH: data collection and clinical trial. VA: manuscript, data analysis. IS: manuscript, data analysis. HE-G: manuscript, data analysis. GS: data collection and clinical trial. JW: oversight, manuscript, data analysis. All authors contributed to the article and approved the submitted version.

FUNDING

The authors above have no relevant financial disclosures or benefits from commercial sources that could create a potential conflict of interest. The entirety of this work was funded by a grant obtained by JA Wilkins through the Canadian Foundation for Innovation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.729681/full#supplementary-material

REFERENCES

1. McInnes IB, Schett G. The Pathogenesis of Rheumatoid Arthritis. *N Engl J Med* (2011) 365:2205–19. doi: 10.1056/NEJMra1004965
2. Smolen JS, Breedveld FC, Burmester GR, Bykerk V, Dougados M, Emery P, et al. Treating Rheumatoid Arthritis to Target: 2014 Update of the Recommendations of an International Task Force. *Ann Rheum Dis* (2016) 75:3–15. doi: 10.1136/annrheumdis-2015-207524
3. Dougados M, Aletaha D, van Riel PL. Disease Activity Measures for Rheumatoid Arthritis. *Clin Exp Rheumatol* (2007) 25:S22–9.
4. Combe B, Rincheval N, Benessiano J, Berenbaum F, Cantagrel A, Daures JP, et al. Five-Year Favorable Outcome of Patients With Early Rheumatoid Arthritis in the 2000s: Data From the ESPOIR Cohort. *J Rheumatol* (2013) 40:1650–7. doi: 10.3899/jrheum.121515
5. Aga AB, Lie E, Uhlig T, Olsen IC, Wierod A, Kalstad S, et al. Time Trends in Disease Activity, Response and Remission Rates in Rheumatoid Arthritis During the Past Decade: Results From the NOR-DMARD Study 2000-2010. *Ann Rheum Dis* (2015) 74:381–8. doi: 10.1136/annrheumdis-2013-204020
6. Galvao TF, Zimmermann IR, da Mota LM, Silva MT, Pereira MG. Withdrawal of Biologic Agents in Rheumatoid Arthritis: A Systematic Review and Meta-Analysis. *Clin Rheumatol* (2016) 35:1659–68. doi: 10.1007/s10067-016-3285-y
7. Ghiti Moghadam M, Vonkeman HE, Ten Klooster PM, Tekstra J, van Schaardenburg D, Starmans-Kool M, et al. Stopping Tumor Necrosis Factor Inhibitor Treatment in Patients With Established Rheumatoid Arthritis in Remission or With Stable Low Disease Activity: A Pragmatic Multicenter, Open-Label Randomized Controlled Trial. *Arthritis Rheumatol* (2016) 68:1810–7. doi: 10.1002/art.39626
8. Smolen JS, Nash P, Durez P, Hall S, Ilivanova E, Irazoque-Palazuelos F, et al. Maintenance, Reduction, or Withdrawal of Etanercept After Treatment With Etanercept and Methotrexate in Patients With Moderate Rheumatoid Arthritis (PRESERVE): A Randomised Controlled Trial. *Lancet* (2013) 381:918–29. doi: 10.1016/S0140-6736(12)61811-X
9. Emery P, Hammoudeh M, FitzGerald O, Combe B, Martin-Mola E, Buch MH, et al. Sustained Remission With Etanercept Tapering in Early Rheumatoid Arthritis. *N Engl J Med* (2014) 371:1781–92. doi: 10.1056/NEJMoa1316133

10. Tweehuysen L, van den Ende C, Beeren F, Been E, van den Hoogen F, den Broeder A. Little Evidence for Usefulness of Biomarkers for Predicting Successful Dose Reduction or Discontinuation of a Biologic Agent in Rheumatoid Arthritis. *Arthritis Rheum* (2017) 69:301–8. doi: 10.1002/art.39946
11. Boja E, Hiltke T, Rivers R, Kinsinger C, Rahbar A, Mesri M, et al. Evolution of Clinical Proteomics and Its Role in Medicine. *J Proteome Res* (2011) 10:66–84. doi: 10.1021/pr100532g
12. Moaddel R, Ubaida-Mohien C, Tanaka T, Lyashkov A, Basisty N, Schilling B, et al. Proteomics in Aging Research: A Roadmap to Clinical, Translational Research. *Aging Cell* (2021) 20:e13325. doi: 10.1111/acel.13325
13. O'Neil LJ, Spicer V, Smolik I, Meng X, Goel RR, Anaparti V, et al. Association of a Serum Protein Signature With Rheumatoid Arthritis Development. *Arthritis Rheumatol* (2021) 73:78–88. doi: 10.1002/art.41483
14. Olivier M, Asmis R, Hawkins GA, Howard TD, Cox LA. The Need for Multi-Omics Biomarker Signatures in Precision Medicine. *Int J Mol Sci* (2019) 20:4781. doi: 10.3390/ijms20194781
15. Haschka J, Englbrecht M, Hueber AJ, Manger B, Kleyer A, Reiser M, et al. Relapse Rates in Patients With Rheumatoid Arthritis in Stable Remission Tapering or Stopping Antirheumatic Therapy: Interim Results From the Prospective Randomised Controlled RETRO Study. *Ann Rheum Dis* (2016) 75:45–51. doi: 10.1136/annrheumdis-2014-206439
16. Rech J, Hueber AJ, Finzel S, Englbrecht M, Haschka J, Manger B, et al. Prediction of Disease Relapses by Multibiomarker Disease Activity and Autoantibody Status in Patients With Rheumatoid Arthritis on Tapering DMARD Treatment. *Ann Rheum Dis* (2016) 75:1637–44. doi: 10.1136/annrheumdis-2015-207900
17. Figueiredo CP, Bang H, Cobra JF, Englbrecht M, Hueber AJ, Haschka J, et al. Antimodified Protein Antibody Response Pattern Influences the Risk for Disease Relapse in Patients With Rheumatoid Arthritis Tapering Disease Modifying Antirheumatic Drugs. *Ann Rheum Dis* (2017) 76:399–407. doi: 10.1136/annrheumdis-2016-209297
18. Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody EN, et al. Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery. *PLoS One* (2010) 5:e15004. doi: 10.1371/journal.pone.0015004
19. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO 3rd, et al. 2010 Rheumatoid Arthritis Classification Criteria: An American College of Rheumatology/European League Against Rheumatism Collaborative Initiative. *Arthritis Rheum* (2010) 62:2569–81. doi: 10.1002/art.27584
20. Felson DT, Smolen JS, Wells G, Zhang B, van Tuyl LH, Funovits J, et al. American College of Rheumatology/European League Against Rheumatism Provisional Definition of Remission in Rheumatoid Arthritis for Clinical Trials. *Ann Rheum Dis* (2011) 70:404–13. doi: 10.1136/ard.2011.149765
21. Tanaka T, Biancotto A, Moaddel R, Moore AZ, Gonzalez-Freire M, Aon MA, et al. Plasma Proteomic Signature of Age in Healthy Humans. *Aging Cell* (2018) 17:e12799. doi: 10.1111/acel.12799
22. Sathyan S, Ayers E, Gao T, Weiss EF, Milman S, Verghese J, et al. Plasma Proteomic Profile of Age, Health Span, and All-Cause Mortality in Older Adults. *Aging Cell* (2020) 19:e13250. doi: 10.1111/acel.13250
23. Govaere O, Cockell S, Tiniakos D, Queen R, Younes R, Vacca M, et al. Transcriptomic Profiling Across the Nonalcoholic Fatty Liver Disease Spectrum Reveals Gene Signatures for Steatohepatitis and Fibrosis. *Sci Transl Med* (2020) 12:eaba4448. doi: 10.1126/scitranslmed.aba4448
24. Lollo B, Steele F, Gold L. Beyond Antibodies: New Affinity Reagents to Unlock the Proteome. *Proteomics* (2014) 14:638–44. doi: 10.1002/pmic.201300187
25. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple Imputation by Chained Equations: What Is it and How Does it Work? *Int J Methods Psychiatr Res* (2011) 20:40–9. doi: 10.1002/mpr.329
26. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res* (2015) 43:e47. doi: 10.1093/nar/gkv007
27. Yu G, Wang LG, Han Y, He QY. clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *OMICS* (2012) 16:284–7. doi: 10.1089/omi.2011.0118
28. Lehallier B, Gate D, Schaum N, Nanasi T, Lee SE, Yousef H, et al. Undulating Changes in Human Plasma Proteome Profiles Across the Lifespan. *Nat Med* (2019) 25:1843–50. doi: 10.1038/s41591-019-0673-2
29. Chen C, Zhang Q, Yu B, Yu Z, Lawrence PJ, Ma Q, et al. Improving Protein-Protein Interactions Prediction Accuracy Using XGBoost Feature Selection and Stacked Ensemble Classifier. *Comput Biol Med* (2020) 123:103899. doi: 10.1016/j.compbiomed.2020.103899
30. Lee TF, Chao PJ, Ting HM, Chang L, Huang YJ, Wu JM, et al. Using Multivariate Regression Model With Least Absolute Shrinkage and Selection Operator (LASSO) to Predict the Incidence of Xerostomia After Intensity-Modulated Radiotherapy for Head and Neck Cancer. *PLoS One* (2014) 9:e89700. doi: 10.1371/journal.pone.0089700
31. Lundberg SM EG, Lee S. Consistent Individualized Feature Attribution for Tree Ensembles. (2018) arXiv2019. https://arxiv.org/abs/1802.03888
32. Felson DT, Smolen JS, Wells G, Zhang B, van Tuyl LH, Funovits J, et al. American College of Rheumatology/European League Against Rheumatism Provisional Definition of Remission in Rheumatoid Arthritis for Clinical Trials. *Arthritis Rheum* (2011) 63:573–86. doi: 10.1002/art.30129
33. Chen S, Blijdorp IC, van Mens LJJ, Bowcutt R, Latuhihin TE, van de Sande MGH, et al. Interleukin 17A and IL-17F Expression and Functional Responses in Rheumatoid Arthritis and Peripheral Spondyloarthritis. *J Rheumatol* (2020) 47:1606–13. doi: 10.3899/jrheum.190571
34. Robert M, Miossec P. IL-17 in Rheumatoid Arthritis and Precision Medicine: From Synovitis Expression to Circulating Bioactive Levels. *Front Med (Lausanne)* (2018) 5:364. doi: 10.3389/fmed.2018.00364
35. Marwa OS, Kalthoum T, Wajih K, Kamel H. Association of IL17A and IL17F Genes With Rheumatoid Arthritis Disease and the Impact of Genetic Polymorphisms on Response to Treatment. *Immunol Lett* (2017) 183:24–36. doi: 10.1016/j.imlet.2017.01.013
36. Taams LS. Interleukin-17 in Rheumatoid Arthritis: Trials and Tribulations. *J Exp Med* (2020) 217:e20192048. doi: 10.1084/jem.20192048
37. Liu R, Lauridsen HM, Amezquita RA, Pierce RW, Jane-Wit D, Fang C, et al. IL-17 Promotes Neutrophil-Mediated Immunity by Activating Microvascular Pericytes and Not Endothelium. *J Immunol* (2016) 197:2400–8. doi: 10.4049/jimmunol.1600138
38. Wallis D. Infection Risk and Biologics: Current Update. *Curr Opin Rheumatol* (2014) 26:404–9. doi: 10.1097/BOR.0000000000000072
39. Maneiro JR, Souto A, Gomez-Reino JJ. Risks of Malignancies Related to Tofacitinib and Biological Drugs in Rheumatoid Arthritis: Systematic Review, Meta-Analysis, and Network Meta-Analysis. *Semin Arthritis Rheum* (2017) 47:149–56. doi: 10.1016/j.semarthrit.2017.02.007
40. Barnabe C, Zheng Y, Ohinmaa A, Crane L, White T, Hemmelgarn B, et al. Effectiveness, Complications, and Costs of Rheumatoid Arthritis Treatment With Biologics in Alberta: Experience of Indigenous and Non-Indigenous Patients. *J Rheumatol* (2018) 10:1344–52. doi: 10.3899/jrheum.170779
41. Subesinghe S, Scott IC. Key Findings From Studies of Methotrexate Tapering and Withdrawal in Rheumatoid Arthritis. *Expert Rev Clin Pharmacol* (2015) 8:751–60. doi: 10.1586/17512433.2015.1077698
42. Smolen JS, Aletaha D. Rheumatoid Arthritis Therapy Reappraisal: Strategies, Opportunities and Challenges. *Nat Rev Rheumatol* (2015) 11:276–89. doi: 10.1038/nrrheum.2015.8
43. Vodencarevic A, Tascilar K, Hartmann F, Reiser M, Hueber AJ, Haschka J, et al. Advanced Machine Learning for Predicting Individual Risk of Flares in Rheumatoid Arthritis Patients Tapering Biologic Drugs. *Arthritis Res Ther* (2021) 23:67. doi: 10.1186/s13075-021-02439-5
44. Norgeot B, Glicksberg BS, Trupin L, Lituiev D, Gianfrancesco M, Oskotsky B, et al. Assessment of a Deep Learning Model Based on Electronic Health Record Data to Forecast Clinical Outcomes in Patients With Rheumatoid Arthritis. *JAMA Netw Open* (2019) 2:e190606. doi: 10.1001/jamanetworkopen.2019.0606
45. Korb A, Pavenstadt H, Pap T. Cell Death in Rheumatoid Arthritis. *Apoptosis* (2009) 14:447–54. doi: 10.1007/s10495-009-0317-y
46. Orange DE, Yao V, Sawicka K, Fak J, Frank MO, Parveen S, et al. RNA Identification of PRIME Cells Predicting Rheumatoid Arthritis Flares. *N Engl J Med* (2020) 383:218–28. doi: 10.1056/NEJMoa2004114
47. Hammer HB, Fagerhol MK, Wien TN, Kvien TK. The Soluble Biomarker Calprotectin (an S100 Protein) Is Associated to Ultrasonographic Synovitis Scores and Is Sensitive to Change in Patients With Rheumatoid Arthritis Treated With Adalimumab. *Arthritis Res Ther* (2011) 13:R178. doi: 10.1186/ar3503

48. Chen YS, Yan W, Geczy CL, Brown MA, Thomas R. Serum Levels of Soluble Receptor for Advanced Glycation End Products and of S100 Proteins Are Associated With Inflammatory, Autoantibody, and Classical Risk Markers of Joint and Vascular Damage in Rheumatoid Arthritis. *Arthritis Res Ther* (2009) 11:R39. doi: 10.1186/ar2645

49. Bach M, Moon J, Moore R, Pan T, Nelson JL, Lood C. A Neutrophil Activation Biomarker Panel in Prognosis and Monitoring of Patients With Rheumatoid Arthritis. *Arthritis Rheumatol* (2020) 72:47–56. doi: 10.1002/art.41062

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

ORIGINAL RESEARCH
published: 19 November 2021
doi: 10.3389/fimmu.2021.775353

B Cell Signatures Distinguish Cutaneous Lupus Erythematosus Subtypes and the Presence of Systemic Disease Activity

Lisa Abernathy-Close[1], Stephanie Lazar[1], Jasmine Stannard[1,2], Lam C. Tsoi[3,4,5], Sean Eddy[6], Syed M. Rizvi[7], Christine M. Yee[7], Emily M. Myers[8], Rajaie Namas[9], Lori Lowe[3,10], Tamra J. Reed[1], Fei Wen[7], Johann E. Gudjonsson[3], J. Michelle Kahlenberg[1] and Celine C. Berthier[6]**

[1] Division of Rheumatology, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, United States, [2] Department of Pediatrics, University of Michigan, Ann Arbor, MI, United States, [3] Department of Dermatology, University of Michigan, Ann Arbor, MI, United States, [4] Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI, United States, [5] Department of Biostatistics, University of Michigan, Ann Arbor, MI, United States, [6] Division of Nephrology, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, United States, [7] Department of Chemical Engineering, University of Michigan, Ann Arbor, MI, United States, [8] Lifebridge Health, Baltimore, MD, United States, [9] Division of Rheumatology, Department of Internal Medicine, Cleveland Clinic Abu Dhabi, Abu Dhabi, United Arab Emirates, [10] Department of Pathology, University of Michigan, Ann Arbor, MI, United States

OPEN ACCESS

Edited by:
Mariele Gatto,
University of Gothenburg, Sweden

Reviewed by:
Ting Gan,
Peking University First Hospital, China
Alex Ortega Loayza,
Oregon Health and Science University, United States

****Correspondence:***
J. Michelle Kahlenberg
mkahlenb@med.umich.edu
Celine C. Berthier
celineb@med.umich.edu

Specialty section:
This article was submitted to Autoimmune and Autoinflammatory Disorders, a section of the journal Frontiers in Immunology

Received: *13 September 2021*
Accepted: *05 November 2021*
Published: *19 November 2021*

Citation:
Abernathy-Close L, Lazar S, Stannard J, Tsoi LC, Eddy S, Rizvi SM, Yee CM, Myers EM, Namas R, Lowe L, Reed TJ, Wen F, Gudjonsson JE, Kahlenberg JM and Berthier CC (2021) B Cell Signatures Distinguish Cutaneous Lupus Erythematosus Subtypes and the Presence of Systemic Disease Activity. Front. Immunol. 12:775353. doi: 10.3389/fimmu.2021.775353

Cutaneous lupus erythematosus (CLE) is a chronic inflammatory skin disease characterized by a diverse cadre of clinical presentations. CLE commonly occurs in patients with systemic lupus erythematosus (SLE), and CLE can also develop in the absence of systemic disease. Although CLE is a complex and heterogeneous disease, several studies have identified common signaling pathways, including those of type I interferons (IFNs), that play a key role in driving cutaneous inflammation across all CLE subsets. However, discriminating factors that drive different phenotypes of skin lesions remain to be determined. Thus, we sought to understand the skin-associated cellular and transcriptional differences in CLE subsets and how the different types of cutaneous inflammation relate to the presence of systemic lupus disease. In this study, we utilized two distinct cohorts comprising a total of 150 CLE lesional biopsies to compare discoid lupus erythematosus (DLE), subacute cutaneous lupus erythematosus (SCLE), and acute cutaneous lupus erythematosus (ACLE) in patients with and without associated SLE. Using an unbiased approach, we demonstrated a CLE subtype-dependent gradient of B cell enrichment in the skin, with DLE lesions harboring a more dominant skin B cell transcriptional signature and enrichment of B cells on immunostaining compared to ACLE and SCLE. Additionally, we observed a significant increase in B cell signatures in the lesional skin from patients with isolated CLE compared with similar lesions from patients with systemic lupus. This trend was driven primarily by differences in the DLE subgroup. Our work thus shows that skin-associated B cell responses distinguish CLE subtypes in patients with and without associated SLE, suggesting that B cell function in skin may be an important link between cutaneous lupus and systemic disease activity.

Keywords: lupus, discoid, B cells, transcriptomic, cutaneous lupus, autoantibodies

INTRODUCTION

Systemic lupus erythematosus (SLE) is complex, chronic, autoimmune disease characterized by hyperreactive B cells and the production of pathogenic autoantibodies (1). SLE involves multiple organ systems, including the skin, where the distinct type of inflammation is termed cutaneous lupus erythematosus (CLE). CLE can occur in isolation or as a skin manifestation associated with underlying systemic lupus erythematosus (SLE) (2). CLE is relatively understudied compared to SLE, which contributes to a lack of understanding of disease heterogeneity in CLE pathogenesis. CLE is a rubric which encompasses clinically and histologically distinct subtypes of CLE: acute cutaneous lupus erythematosus (ACLE), subacute cutaneous lupus erythematosus (SCLE), or chronic lupus erythematosus (CCLE), with discoid lupus erythematosus (DLE) being the most common subtype (3–5). While there are consistently observed cellular and molecular features in patients with CLE and/or SLE, such as a type I interferon (IFN) gene signature in the blood and skin (6–10) and peripheral B cell dysfunction (11, 12), the shared and unique molecular and cellular features of ACLE, SCLE, and CCLE remain poorly understood. Indeed, basic transcriptional comparisons have not identified robust distinguishing molecular signatures between subtypes (13, 14). Further, DLE is more likely to occur without underlying SLE compared to ACLE or SCLE (2, 15), yet it is not clear if the presence or absence of systemic disease is related to the differences observed in cutaneous manifestations of lupus (16).

In this study, we sought to investigate cellular and transcriptional differences in lesional skin biopsies across CLE subtypes, including DLE, ACLE, and SCLE, and explore how cutaneous lesional immunophenotypes relates to systemic lupus disease. Using novel analyses, we found that DLE lesions harbor a unique immunoglobulin signature and an enrichment of skin B cells compared to ACLE or SCLE lesions. Intriguingly, this B cell signature was highest in patients with CLE without concomitant SLE, including within the entire cohort of DLE patients (2). Our results demonstrate that a B cell gene signature in the skin distinguishes DLE from ACLE and SCLE and that increased B cells in the skin of DLE patients is indicative of a lower rate of accompanying systemic disease. These data suggest that B cell transcriptional programs are more activated in DLE lesions relative to ACLE or SCLE lesions and may play a role in immunopathogenesis divergence across CLE subtypes. This work supports future exploration of utilizing a skin B cell score as a clinical marker of SLE risk, especially in DLE patients.

MATERIALS AND METHODS

Study Design

We applied a tissue transcriptome-driven sequential analysis strategy. Gene expression profiles from 90 cases of CLE that include DLE (n=47) and SCLE (n=43), as well as 13 healthy control skin biopsies were used as a discovery cohort (13). Subsequent profiles of 60 skin biopsies that include DLE (n=20), SCLE (n=20) and ACLE (n=20) served as a validation cohort along with 4 additional healthy control skin biopsies. The details of the discovery cohort and sample collection protocol are previously described (13). In brief, skin biopsies were identified *via* a SNOMED search of the University of Michigan Pathology Database using the search terms "lupus" and "cutaneous lupus". Patients who met both clinical and histologic criteria for DLE or SCLE or ACLE were included in the study. Patients with drug-induced CLE were excluded from this study. The average time from diagnosis to skin biopsy ranged from two to four years for patients in the discovery cohort. Gender-, age- and race-matched healthy controls were identified and utilized for studies that compared CLE to normal healthy control skin (n=13 in the discovery cohort, n=4 in the validation cohort). Clinical data information can be found in **Supplementary Table 1**. Systemic disease was defined by ACR criteria ≥4 (17).

Gene Expression Analysis

For both discovery and validation cohorts, transcriptome analysis was performed on skin biopsies using Affymetrix ST2.1 GeneChips as previously published (13). Data processing details for the discovery cohort can be found in Berthier et al. (13). In brief, normalized expression data were log2-transformed and batch-corrected. FDR was applied to account for multiple testing. The CEL-files and processed data are available at Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) under the reference number GSE81071 (discovery cohort) and GSE184989 (validation cohort).

For the validation cohort, samples were processed and normalized using the RMA approach (18), and average expression was taken if more than one probesets mapping to a gene. We then estimated and controlled the latent confounding variables (19) for the limma-based differential expression analysis (20, 21).

IFN Score Calculation

For both discovery and validation cohorts, IFN score was calculated from the gene expression data as previously described (13).

Weighted Gene Co-Expression Network Analysis

Weighted gene co-expression network analysis (WGCNA) (22) was performed on the 20,410 genes of the discovery cohort. Briefly, WGCNA was used to aggregate genes into co-expression modules representing gene expression patterns across all patient samples. Co-expression modules were named and represented by a unique color. A module eigenene value (the first principal component of each module gene set) was generated in each sample used as representation of the module. Each module eigengene value was associated with available clinical variables: Cutaneous Lupus Erythematosus Disease Area and Severity Index (CLASI), Systemic Lupus Erythematosus Disease Activity Index (SLEDAI), systemic versus non-systemic disease status, DLE versus SCLE status, and IFN score.

Heatmap Generation, Cell Type Enrichment and Literature-Based Pathway Analyses

Heatmaps of gene expression datasets were generated using the Morpheus software (https://software.broadinstitute.org/morpheus). Cell type enrichment analysis was performed as previously reported (13) on the normalized datasets of 20,410 genes (discovery cohort) and of 29,405 genes (validation cohort) using the xCell webtool (http://xcell.ucsf.edu/) (23). Canonical pathways were identified using Ingenuity Pathway Analysis software (IPA) (www.ingenuity.com).

Tissue CyTOF

Formalin-fixed, paraffin-embedded (FFPE) skin biopsy tissue sections from lesional skin of patients with ACLE, SCLE, or DLE were analyzed using the Hyperion imaging CyTOF system (Fluidigm) as previously described (24) with modifications of the antibody panel. Specifically, metal-tagged antibodies including pan-keratin (C11, Biolegend), BDCA2 (Polyclonal, R&D Systems), CD56 (123C3, ThermoFisher Scientific), HLA-DR (LN3, Biolegend), CD11c (EP1347Y, Abcam), and CD4 (EPR6855, Fluidigm) were added in this study.

Immunohistochemistry

CLE skin biopsies from lesions of patients with ACLE, SCLE, or DLE as well as healthy controls were collected and fixed in formalin. Formalin-fixed, paraffin-embedded skin biopsy sections were assayed by chromogenic immunostaining for pan-leukocytes (anti-CD45, HI30, eBioscience), B cells (anti-CD20, L26, Abcam) and memory B cells (anti-CD27, polyclonal, R&D Systems). Antigen retrieval was achieved by heating sections in sodium citrate buffer (pH 6.0) prior to antibody incubation. A minimum of 3 patients per disease status group were assayed and representative images are shown.

Statistical Analyses

Statistical analysis of clinical data and gene score comparisons were generated using an unpaired parametric t-test with GraphPad Prism software version 8.0.0; p-values<0.05 were considered statistically significant and reported in all Figures. All comparisons across all groups were performed; for clarity, only the most relevant were reported if statistically significant.

RESULTS

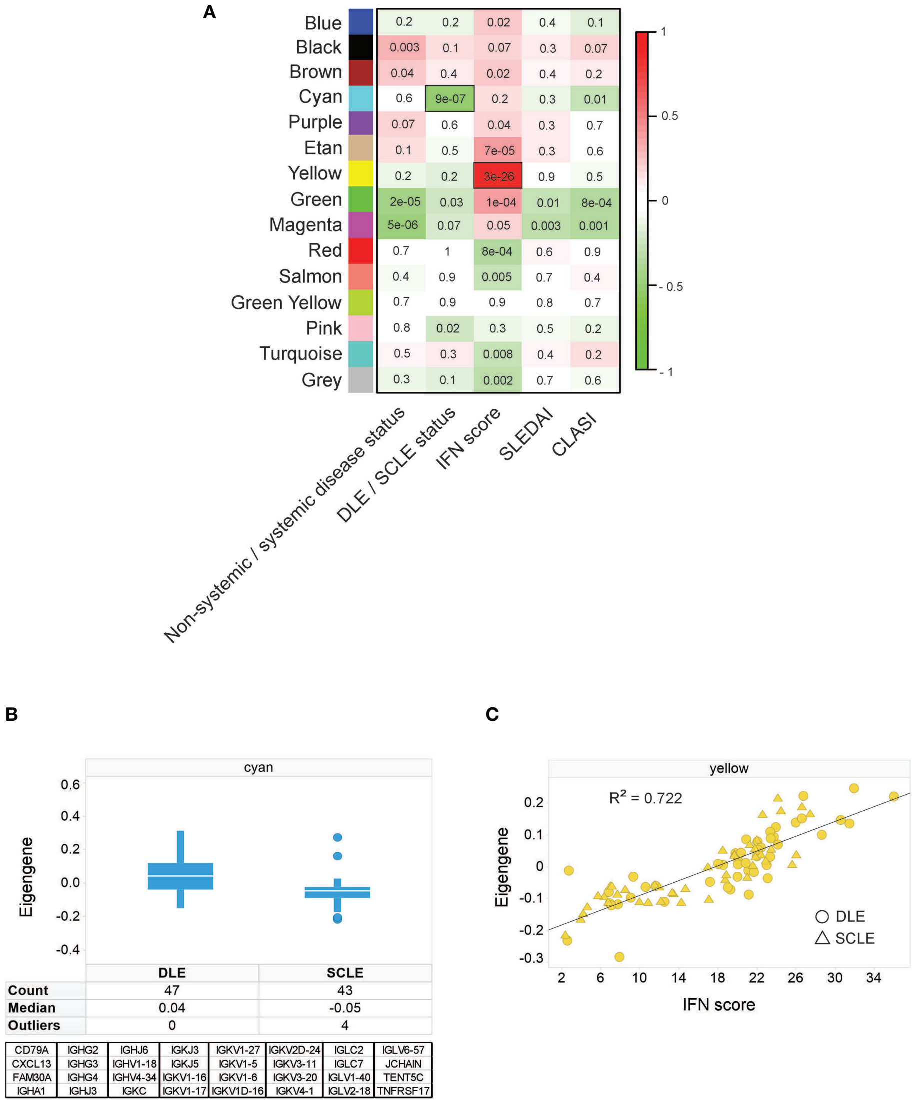
DLE Lesions Harbor a Unique Skin Immunoglobulin Gene Signature Compared to SCLE Lesions

To identify unique features amongst CLE subtypes, weighted gene correlation network analysis (WGCNA) was performed on an initial discovery skin cohort to identify modules of genes correlating with available patient clinical variables: CLASI, SLEDAI, systemic versus non-systemic disease status, DLE versus SCLE status, and IFN score. The resulting modules were categorized by color and correlations of each module eigengene with clinical variables depicted in the module-trait relationship heatmap (**Figure 1A**). This analysis identified that the cyan module was one of the modules with the strongest correlation with clinical variables and was significantly higher in DLE compared to SCLE status (**Figure 1B**). This cyan module was775353 composed of 32 genes, 26 of which were immunoglobulin genes (**Figure 1B**). This result was confirmed in a separate validation cohort that also included patients with ACLE (**Supplementary Figure 1A**). Further probing by Ingenuity pathway analysis also revealed significant enrichment for several B cell-related pathways (**Supplementary Table 2**) including B cell receptor signaling ($p=6.31x10^{-33}$), B cell signaling pathway ($p=1.58x10^{-31}$), B cell activating factor signaling ($p=4.79x10^{-02}$), and B cell development ($p=4.79x10^{-02}$). The yellow module from the WGCNA analysis, represented by 746 genes (**Supplementary Table 3),** was significantly correlated with the increased IFN score in both the CLE discovery cohort (**Figure 1C**, $r^2 = 0.722$, $p=3x10^{-28}$) and validation cohort (**Supplementary Figure 1B**, $r^2 = 0.832$, p<0.0001). These data demonstrate that a stronger immunoglobulin gene signature was observed in lesional skin from patients with DLE when compared to lesions from those with SCLE, and that a high skin IFN score was correlated with active CLE lesions, regardless of cutaneous disease subtype.

DLE Lesions Are Associated With a Higher Skin B Cell Signature Compared to ACLE or SCLE

To explore whether the skewed cutaneous immunoglobulin gene signature detected in DLE lesions coincided with an increase in skin B cell subsets compared to other CLE subtypes, we utilized the xCell algorithm which performs cell type enrichment analysis from tissue gene expression profiles (23). The heterogeneous cellular landscape of tissue expression profiles can be evaluated with the xCell enrichment scores generated for each cell type. We performed this analysis on both the discovery and the validation cohorts which included normalized gene expression from the skin of healthy controls or lesions from patients with DLE, SCLE, or ACLE. (**Supplementary Figure 2** and **Supplementary Table 4**).

The cell type enrichment analysis showed enrichment for B cell subsets in patients with DLE compared to SCLE or ACLE skin lesions (**Figure 2A**). Specifically, cell type enrichment analysis from both the discovery and validation cohorts revealed significantly higher gene expression signatures for B cells (p=0.0001 and p<0.0001, respectively), naïve B cells (p=0.0011 and p<0.0001, respectively) and memory B cells (p<0.0001 and p=0.0001, respectively) in skin lesions from patients with DLE compared to SCLE (**Figure 2B**). Furthermore, B cell enrichment scores of B cells, naïve B cells, and memory B cells were all significantly lower in ACLE lesions compared to DLE (**Figure 2B**, p=0.0025, p=0.0007, p=0.0070, respectively). No significant difference in the enrichment of B cell subsets between lesions from patients with SCLE and ACLE was detected by cell type enrichment analysis (**Figure 2B**). These results show that when CLE is examined according to cutaneous disease subtype, a significant enrichment in overall and subsets of skin B cell gene expression programs is detected in DLE lesions compared to SCLE or ACLE lesions.



	Non-systemic / systemic disease status	DLE / SCLE status	IFN score	SLEDAI	CLASI
Blue	0.2	0.2	0.02	0.4	0.1
Black	0.003	0.1	0.07	0.3	0.07
Brown	0.04	0.4	0.02	0.4	0.2
Cyan	0.6	9e-07	0.2	0.3	0.01
Purple	0.07	0.6	0.04	0.3	0.7
Etan	0.1	0.5	7e-05	0.3	0.6
Yellow	0.2	0.2	3e-26	0.9	0.5
Green	2e-05	0.03	1e-04	0.01	8e-04
Magenta	5e-06	0.07	0.05	0.003	0.001
Red	0.7	1	8e-04	0.6	0.9
Salmon	0.4	0.9	0.005	0.7	0.4
Green Yellow	0.7	0.9	0.9	0.8	0.7
Pink	0.8	0.02	0.3	0.5	0.2
Turquoise	0.5	0.3	0.008	0.4	0.2
Grey	0.3	0.1	0.002	0.7	0.6

	DLE	SCLE
Count	47	43
Median	0.04	-0.05
Outliers	0	4

CD79A	IGHG2	IGHJ6	IGKJ3	IGKV1-27	IGKV2D-24	IGLC2	IGLV6-57
CXCL13	IGHG3	IGHV1-18	IGKJ5	IGKV1-5	IGKV3-11	IGLC7	JCHAIN
FAM30A	IGHG4	IGHV4-34	IGKV1-16	IGKV1-6	IGKV3-20	IGLV1-40	TENT5C
IGHA1	IGHJ3	IGKC	IGKV1-17	IGKV1D-16	IGKV4-1	IGLV2-18	TNFRSF17

FIGURE 1 | Weighted gene correlation network analysis identifies an immunoglobulin signature associated with DLE skin lesion status, independent of IFN score. **(A)** Module-trait relationship heatmap. Each module eigengene was correlated with the indicated clinical parameter. For categorical parameters, non-systemic/ systemic disease and DLE/SCLE, numerical values were assigned to each categorical group. The scale bar on the right represents the correlation coefficient with green for negative correlation and red for positive correlation, p-values for each correlation are presented on the heatmap. The yellow module was the module with the strongest positive correlation with IFN score and the cyan module was the module that had the strongest negative correlation with the DLE versus SCLE lesion status. **(B)** The cyan module eigengene from the SCLE and DLE lesions encompasses a 32-gene, primarily immunoglobulin signature. **(C)** The yellow module eigengene was significantly correlated with IFN score (r = 0.85, p = 3E-28). The data in each panel represent 47 DLE patients and 43 SCLE patients.

Skin B Cell Enrichment in CLE Is Associated With Non-Systemic Lupus and Discoid Lesions

We then sought to determine if the B cell signature detected in CLE lesions was associated with systemic disease status, including within a particular CLE subtype. For this, we first grouped all CLE lesions together and analyzed them based on whether the patient had systemic SLE or CLE without systemic disease. Intriguingly, we identified a significant difference in lesional B cell subsets vs. healthy control only in patients with CLE without associated SLE (p=0.0003). No increase in B cell subsets were noted when lesions from SLE patients were compared with healthy controls (**Figure 3A**). This trend was less when B cells were subsetted into naïve and memory, but in all instances, CLE lesions taken from patients without SLE exhibited significantly more B cell-associated gene expression. (**Figure 3A**). Because we observed differences in the lesional skin B cell enrichment score in patients with CLE based on SLE status, we further explored the impact of CLE subtype on this finding. There were significantly higher total B cell enrichment scores in

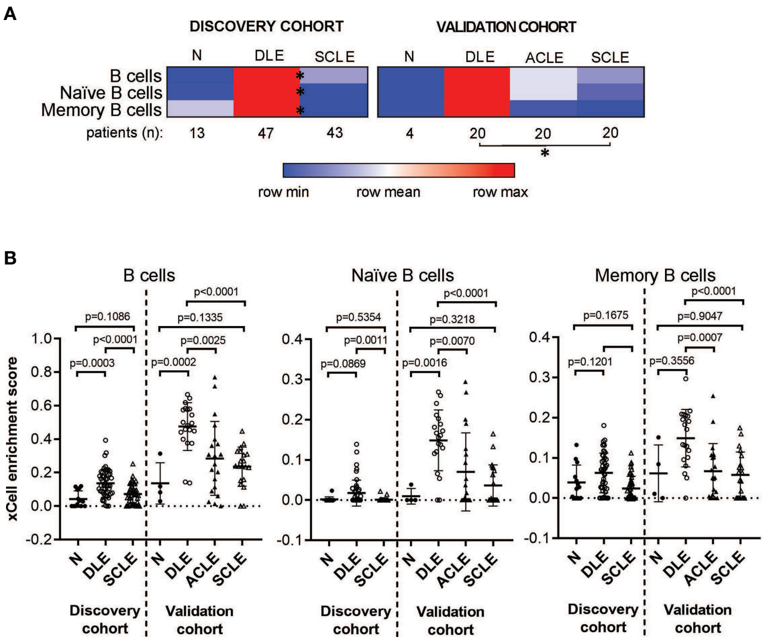


FIGURE 2 | Cell type enrichment analysis using xCell tool reveals a B cell signature higher in DLE lesional skin compared to SCLE and ACLE lesions. **(A)** Heatmap of the B cell subtypes representing average xCell score for each skin lesion type compared to normal healthy controls (N). *p-value < 0.05 in SCLE versus DLE. **(B)** xCell enrichment score for B cells, naïve B cells and memory B cells in lesional skin from patients with DLE, ACLE, SCLE as well as normal healthy controls (N) in both the discovery and the validation cohort. Comparisons were made *via* unpaired Students' t-test.

lesions from patients with isolated DLE compared to DLE with underlying SLE (p=0.0008). A similar trend was seen for SCLE but this did not reach significance (p=0.077) (**Figure 3B**). This B cell enrichment was also specific to non-systemic DLE versus non-systemic SCLE when total B cells (p=0.0001), naïve B cells (p=0.0489), and memory B cells (p=0.0183) were compared (**Figure 3**). Thus, lack of associated SLE is associated with increased B cell signatures in CLE lesions, and this holds true even within the DLE subtype when patients with and without associated SLE are compared.

Peripheral Autoantibody Status Is Related to the Degree of B Cell Enrichment in Cutaneous Lupus Lesions

Given that our data show that a prominent skin B cell gene signature is most pronounced in DLE lesions with a lack of underlying SLE, we sought to explore whether there is a relationship between skin B cell enrichment in lesional skin and the presence of peripheral lupus autoantibodies. We thus examined the lesional skin xCell B cell enrichment scores of DLE or SCLE patients subsetted by the presence of absence of key diagnostic lupus autoantibodies at the time of skin biopsy (**Figure 4**). As expected, DLE or SCLE patients with systemic disease were overall more likely to test positive for lupus antibodies than patients without systemic disease (**Figure 4**). Surprisingly, however, anti-nuclear antibody (ANA) negative DLE patients still demonstrated elevations in their cutaneous B cell signatures when compared to ANA+ DLE patients (**Figure 4A**). This was also true when the comparisons were made between anti-dsDNA- versus anti-dsDNA+ patients (**Figure 4A**). No differences were noted between ANA or anti-dsDNA positivity and B cell signatures in SCLE biopsies (**Figure 4A**). Interestingly, there was no difference between B cell enrichment score in DLE lesional skin for anti-Smith, anti-Ro, and anti-phospholipid antibodies (**Figures 4B, C**). In SCLE patients, only those positive for anti-Smith antibodies had significantly lower lesional skin B cell enrichment scores (**Figure 4B**, p=0.0235). Furthermore, we observed that skin B cell enrichment scores remained significantly higher in DLE lesions compared to SCLE lesions among patients who tested negative for ANA (p=0.0010) or anti-dsDNA (p=0.0009) (**Figure 4A**), anti-Smith (p=0.0059) or anti-Ro (p=0.0131) (**Figure 4B**), or anti-phospholipid antibodies (p=0.0066) (**Figure 4C**). Taken together, these data show that patients with DLE have elevated cutaneous B cell signatures without a concurrent subsequent rise in peripheral autoantibodies. This suggests that the function of B cells in CLE lesions may go beyond generation of antibody secreting cells or that cutaneously-produced antibodies either do not reach circulation or are against antigens that are not tested for on routine clinical testing.

DLE Lesions Exhibit Higher Skin B Cell Numbers Compared to SCLE or ACLE

To validate the transcriptional data and to expand our investigation to ACLE, immune cell populations were

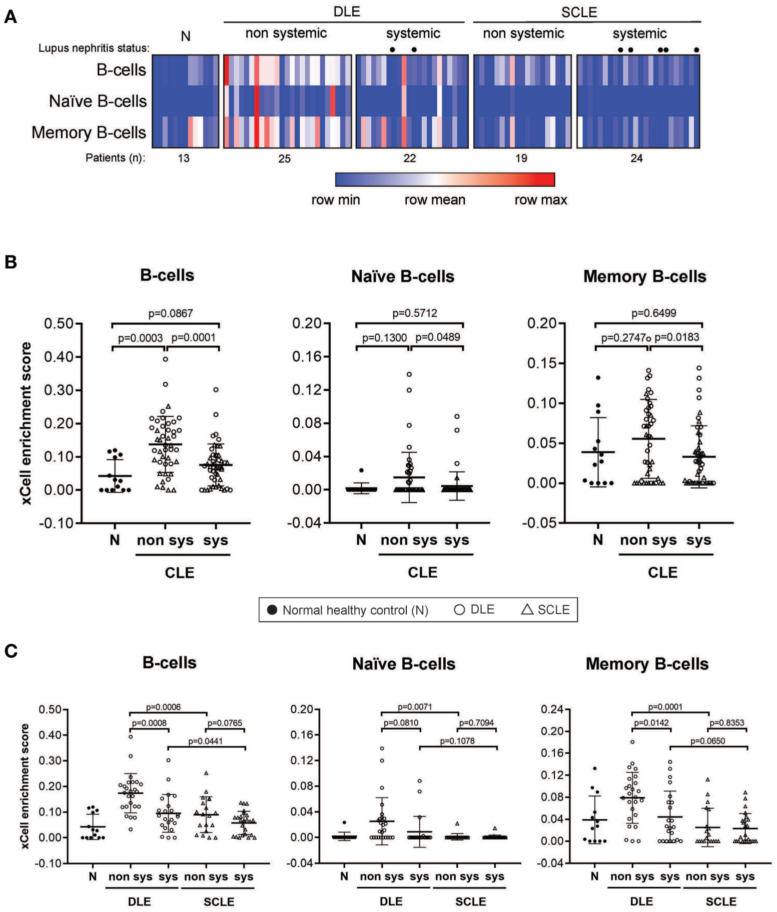


FIGURE 3 | B cell subset enrichment score in lesional skin from CLE patients with and without systemic lupus. **(A)** Heatmap of the B cell subtypes representing the xCell enrichment score for each patient from the discovery cohort. **(B)** xCell enrichment score for B cell subtypes in normal healthy controls (N) (n = 13) and all CLE patients with and without systemic lupus (n = 46 and n = 44, respectively). **(C)** xCell enrichment score for B cell subtypes in DLE patients with and without systemic lupus (n = 22 and n = 25, respectively) and SCLE patients with and without systemic lupus (n = 24 and n = 19, respectively). Comparisons were made *via* unpaired Students' t-test.

enumerated in lesional skin biopsies from patients with DLE, SCLE, or ACLE lesions by tissue CyTOF using a 16-antibody panel. While most immune cell populations, except for natural killer (NK) cells, neutrophils, and plasmablasts, were detectible in CLE lesions (**Supplementary Figure 3**), increased B cell numbers were only seen in DLE and ACLE lesions but not in SCLE (**Figure 5A**). This coincides with the skewing of certain B cell related genes in DLE>ACLE>SCLE lesions (**Figures 5B, C**).

We then sought to confirm the distinct differences of a skin B cell signature across CLE subtypes *via* immunohistochemistry. Skin sections from ACLE, SCLE, and DLE lesions were probed with a pan-leukocyte marker (CD45), a broad B cell marker (CD20), and a marker for mature B cells (CD27) by immunohistochemistry (**Figure 6**). While an increase in CD45+ immune cells was observed in the skin of patents with CLE, we observed the highest infiltration of CD20+ B cells and CD27+ mature B cells in DLE lesional skin, followed by ACLE lesions, and SCLE lesional skin harbored the lowest infiltration of these immune cells among these CLE subtypes (**Figure 6** and **Supplementary Figure 4**). Taken together, these data reveal a gradient of B cell numbers and associated B cell marker gene expression with the highest in DLE and the lowest in SCLE.

DISCUSSION

The cellular and molecular basis of disease heterogeneity in CLE and the variability in which systemic lupus erythematosus (SLE) occurs in patients with different CLE subtypes remain important research objectives to understand lupus pathogenesis and develop precision therapies. To that end, we explored the transcriptional and cellular phenotypes of skin biopsies from patients with ACLE, SCLE, and DLE. As expected, we observed that interferon (IFN) genes were globally upregulated in both DLE and SCLE and therefore did not discriminate between these subtypes of disease. However, we subsequently identified a B cell

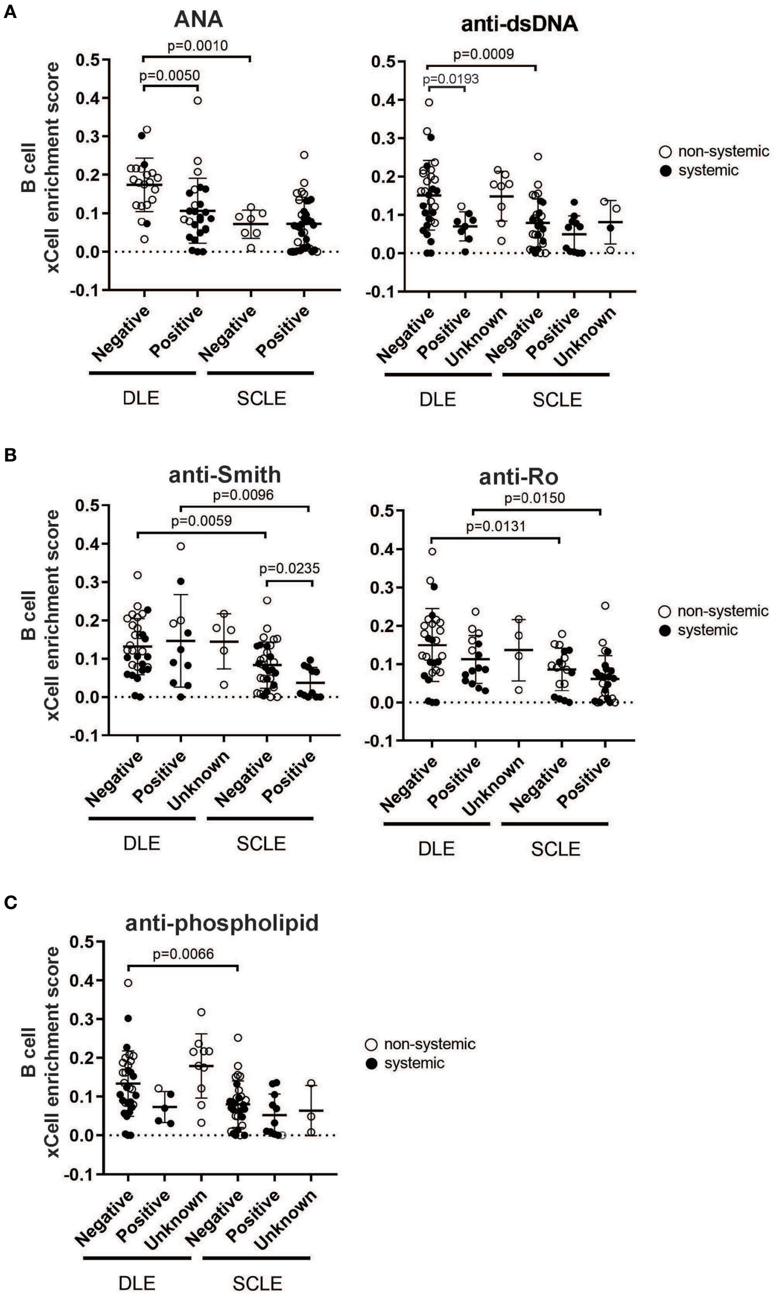


FIGURE 4 | The relationship between skin B cell enrichment and the presence of circulating SLE autoantibodies in patients with DLE or SCLE. Patients with active DLE or SCLE skin lesions were stratified by the presence (positive) or absence (negative) of SLE autoantibodies at the time of biopsy. The patients in which the status of a particular autoantibody was not known at the time of biopsy were classified as "unknown". **(A)** ANA and anti-dsDNA. **(B)** Anti-Smith and anti-Ro. **(C)** Anti-phospholipid. Comparisons were made *via* unpaired Students' t-test.

gene signature in CLE skin that did indeed distinguish DLE from ACLE and SCLE, and this CLE-associated gene signature was highest in DLE lesions without associated systemic disease. These data indicate that while type I IFNs are known to contribute to the recruitment and activation of B cells in autoimmune disease (25–27), they may not be critical drivers in the differential recruitment of B cells observed in DLE skin.

Autoimmune responses and the contribution of B cells in SLE pathogenesis are well described, yet a role for skin-associated B cells in CLE is less apparent. There is considerable interest in the development of murine models to explore the contribution of B cells and other immune cell populations in the skin to cutaneous lupus pathogenesis (28, 29) and studies in the role of B cells in CLE patients are emerging. Indeed, our data suggest that while

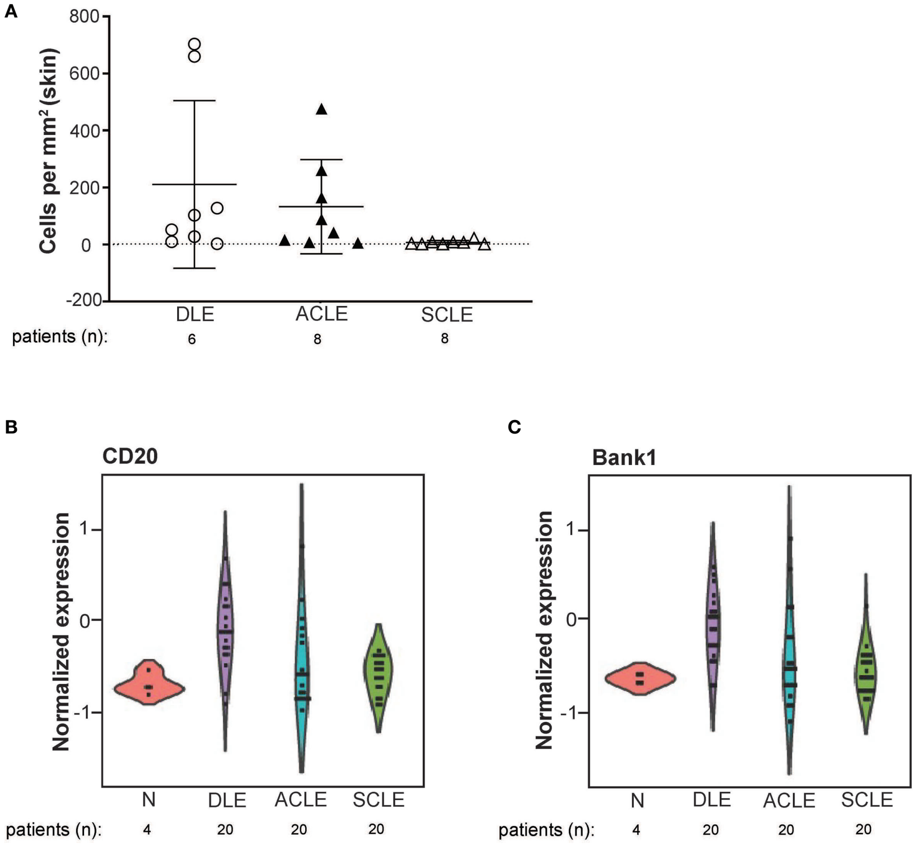


FIGURE 5 | B cell quantification by tissue CyTOF and gene expression in lesional skin from ACLE, SCLE, and DLE patients. **(A)** The number of B cells numbers per millimeter of skin were quantified in SCLE (n = 8), ACLE (n = 8), and DLE (n = 8) lesions. Normalized gene expression of **(B)** CD20, and **(C)** Bank1 from SCLE (n = 20), ACLE (n = 20), and DLE (n = 20) lesional skin and normal healthy control skin (N) (n = 4).

DLE is more likely to occur without underlying SLE compared to ACLE or SCLE (2, 5, 15), we detected a higher B cell signature in DLE lesions in patients without SLE that does not reflect peripheral autoantibody status. Our data mirror a previous study by Magro et al. that reported robust CD20 staining in 14/18 DLE lesions and only moderate CD20 staining in 9 SLE lesional biopsies (classification of biopsy subtype was not provided for SLE patients) (30). A more recent study that explored peripheral B cells in CCLE patients found that patients that lack systemic disease share peripheral B cell abnormalities with SLE patients (11). However, tissue-specific B cell responses in the skin of patients were not examined.

Thus, our data suggest that understanding tissue-specific B cell responses may be important for disease phenotype and possibly for predicting medication responses. Indeed, some small studies have suggested that SLE patients with refractory DLE respond better than SLE patients with SCLE to B cell depleting therapy (31). Another large study reported on 82 patients with SLE all of whom were treated with Rituximab (32). Importantly, no CLE without SLE patients were in this study. In this analysis, no DLE+SLE patients responded to Rituximab, yet of the ACLE patients that responded, negative anti-RNP and negative anti-Ro antibodies were associated with better response (32). Thus, we would propose that based on our data, DLE patients without SLE, especially those without a positive ANA, may be an important patient group to study for the effects of B cell depletion. Further clinical studies should address the benefit of B cell targeted therapy in isolated, refractory DLE.

Our study has several limitations. First, our study was performed retrospectively on archived patient samples. While this allows us to analyze a larger number of samples, we are limited in the clinical data and long-term follow-up that we were able to collect. Secondly, our discovery cohort did not include ACLE patients secondary to the design of that initial work. Future work should explore changes in skin B cell signatures over time in longitudinally collected patient samples and should include additional phenotyping studies to understand the role of B cells in the skin, whether they are contributing to antibody secretion, and whether their depletion can be a viable therapy in the right subset of patients.

In summary, we have identified a transcriptional B cell signature that is highest in DLE>ACLE>SCLE patients and that is most prominent when the CLE lesions occur without

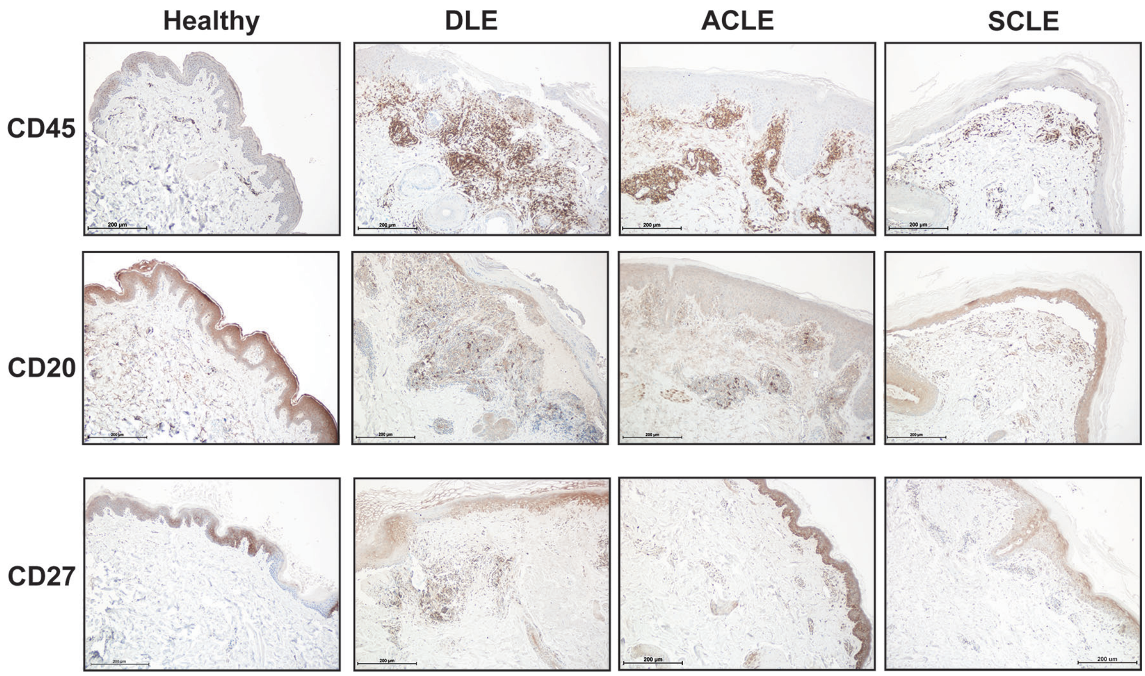


FIGURE 6 | Immunohistochemistry staining for total immune cells and B cell subsets in lesional skin from healthy controls or patients with DLE, ACLE, or SCLE. Formalin-fixed paraffin embedded tissue sections from skin were stained for CD45+ total leukocytes, CD20+ B cells, and CD27+ mature B cells. Representative images from 3-5 patients of each subtype are shown at 100X magnification with a scale bar of 200 μm.

associated SLE. This was validated by immunostaining for both naïve and memory B cell populations in lesional skin. Interestingly, patients with skin lesions and positive autoantibodies tend to have a lower B cell enrichment score in the skin. This data has important implications for trial design for patients with isolated CLE, as treatment options for refractory CLE without SLE are limited. Further study into the role of B cells, the recruitment and differentiation in lesional skin, and the types of antibody secreting cells present will further enhance our ability to diagnose and treat CLE.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: https://www.ncbi.nlm.nih.gov/geo/ under the accession numbers GSE184989 and GSE81071 (https://www.ncbi.nlm.nih.gov/geo/).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by IRBMED at University of Michigan. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

Design of the study was performed by LA-C, JMK, JG, and CB. Sample and clinical data collection was performed by SL, RN, JS, and EM. Conducting experiments and data acquirement were done by LA-C, SR, CY, RN, LL, TR, and FW. Data and bioinformatics analyses were performed by LA-C, LT, SE, and CB. And interpretation of data and writing of drafts were performed by LA-C, LT, SE, JG, JMK, and CB. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the National Institute of Arthritis and Musculoskeletal and Skin Diseases through awards R03AR072107, K24AR076975, and R01AR071384 (to JMK), and P30AR079374, and R01 AI130025 (to JG). Additional support was received from the Taubman Institute Innovative Projects Program (JMK, JG, and FW), and Dermatology Foundation (to JG). Some data in the validation cohort was also collected through support from Celgene to JMK and JG. Tissue CyTOF work was supported by NIH Office of the Director through award S10-OD020053 (to FW) and National Cancer Institute through award P30-CA046592 (to FW), and National Science Foundation through award 1653611 (to FW). In addition, LA-C was supported by grant number UL1TR002240 from the National Center for Advancing

Translational Sciences (NCATS). We thank the support of George M. O'Brien Michigan Kidney Translational Research Core Center (P30DK081943).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.775353/full#supplementary-material

Supplementary Figure 1 | Weighted correlation network analysis of DLE versus SCLE status and associated cyan and the IFN-score associated yellow modules in the validation cohort. **(A)** A total of 27 of the 32 Cyan module associated genes from the discovery cohort were also significantly higher in DLE compared to SCLE and ACLE, confirming the association of the cyan module with the DLE versus SCLE status (n = 20 per CLE subtype). **(B)** A total of 559 from the 746 yellow module associated genes from the discovery cohort also significantly correlated with the IFN score in the validation cohort (p-value < 0.0001).

Supplementary Figure 2 | Cell type enrichment analysis using xCell tool, on both the discovery and the validation cohorts. Heatmap of relevant cell types representing average xCell score for each CLE disease compared to controls. Comparisons were made *via* unpaired Students' t-test. Detailed summary statistics are presented in **Supplementary Table 4**. The asterisks represent the statistically significant changes in DLE compared to SCLE or ACLE compared to SCLE (p-value < 0.05).

Supplementary Figure 3 | Immune cell quantification by tissue CyTOF and gene expression in lesional skin from ACLE, SCLE, and DLE patients. Immune cell populations in the skin were quantified in SCLE, ACLE, and DLE lesions.

Supplementary Figure 4 | Immunohistochemistry staining for B cell subsets in lesional skin from additional patients with DLE, ACLE, or SCLE. Formalin-fixed paraffin embedded tissue sections from skin were stained for CD20+ B cells and CD27+ mature B cells (n = 3 patients per CLE subtype). Representative images are shown at 100X magnification with a scale bar of 200 μm.

Supplementary Table 2 | Ingenuity pathway analysis from the 32 WGCNA cyan module genes: regulated canonical pathways. B cell-related pathways are highlighted in bold.

REFERENCES

1. Lipsky PE. Systemic Lupus Erythematosus: An Autoimmune Disease of B Cell Hyperactivity. *Nat Immunol* (2001) 2(9):764–6. doi: 10.1038/ni0901-764
2. Garelli CJ, Refat MA, Nanaware PP, Ramirez-Ortiz ZG, Rashighi M, Richmond JM. Current Insights in Cutaneous Lupus Erythematosus Immunopathogenesis. *Front Immunol* (2020) 11:1353. doi: 10.3389/fimmu.2020.01353
3. Stannard JN, Kahlenberg JM. Cutaneous Lupus Erythematosus: Updates on Pathogenesis and Associations With Systemic Lupus. *Curr Opin Rheumatol* (2016) 28(5):453–9. doi: 10.1097/BOR.0000000000000308
4. Okon LG, Werth VP. Cutaneous Lupus Erythematosus: Diagnosis and Treatment. *Best Pract Res Clin Rheumatol* (2013) 27(3):391–404. doi: 10.1016/j.berh.2013.07.008
5. Grönhagen CM, Fored CM, Granath F, Nyberg F. Cutaneous Lupus Erythematosus and the Association With Systemic Lupus Erythematosus: A Population-Based Cohort of 1088 Patients in Sweden. *Br J Dermatol* (2011) 164(6):1335–41. doi: 10.1111/j.1365-2133.2011.10272.x
6. Baechler EC, Batliwalla FM, Karypis G, Gaffney PM, Ortmann WA, Espe KJ, et al. Interferon-Inducible Gene Expression Signature in Peripheral Blood Cells of Patients With Severe Lupus. *Proc Natl Acad Sci U S A* (2003) 100 (5):2610–5. doi: 10.1073/pnas.0337679100
7. Bennett L, Palucka AK, Arce E, Cantrell V, Borvak J, Banchereau J, et al. Interferon and Granulopoiesis Signatures in Systemic Lupus Erythematosus Blood. *J Exp Med* (2003) 197(6):711–23. doi: 10.1084/jem.20021553
8. Crow MK. Type I Interferon in the Pathogenesis of Lupus. *J Immunol* (2014) 192(12):5459–68. doi: 10.4049/jimmunol.1002795
9. Sarkar MK, Hile GA, Tsoi LC, Xing X, Liu J, Liang Y, et al. Photosensitivity and Type I IFN Responses in Cutaneous Lupus are Driven by Epidermal-Derived Interferon Kappa. *Ann Rheum Dis* (2018) 77(11):1653–64. doi: 10.1136/annrheumdis-2018-213197
10. Zhu JL, Tran LT, Smith M, Zheng F, Cai L, James JA, et al. Modular Gene Analysis Reveals Distinct Molecular Signatures for Subsets of Patients With Cutaneous Lupus Erythematosus. *Br J Dermatol* (2021) 185(3):563–72. doi: 10.1111/bjd.19800
11. Jenks SA, Wei C, Bugrovsky R, Hill A, Wang X, Rossi FM, et al. B Cell Subset Composition Segments Clinically and Serologically Distinct Groups in Chronic Cutaneous Lupus Erythematosus. *Ann Rheum Dis* (2021). doi: 10.1136/annrheumdis-2021-220349
12. Kil LP, Hendriks RW. Aberrant B Cell Selection and Activation in Systemic Lupus Erythematosus. *Int Rev Immunol* (2013) 32(4):445–70. doi: 10.3109/08830185.2013.786712
13. Berthier CC, Tsoi LC, Reed TJ, Stannard JN, Myers EM, Namas R, et al. Molecular Profiling of Cutaneous Lupus Lesions Identifies Subgroups Distinct From Clinical Phenotypes. *J Clin Med* (2019) 8(8). doi: 10.3390/jcm8081244
14. Ko WC, Li L, Young TR, McLean-Mandell RE, Deng AC, Vanguri VK, et al. Gene Expression Profiling in Skin Reveals Strong Similarities Between Subacute and Chronic Cutaneous Lupus That are Distinct From Lupus Nephritis. *J Invest Dermatol* (2021). doi: 10.1016/j.jid.2021.04.030
15. Vera-Recabarren MA, García-Carrasco M, Ramos-Casals M, Herrero C. Comparative Analysis of Subacute Cutaneous Lupus Erythematosus and Chronic Cutaneous Lupus Erythematosus: Clinical and Immunological Study of 270 Patients. *Br J Dermatol* (2010) 162(1):91–101. doi: 10.1111/j.1365-2133.2009.09472.x
16. Maz MP, Michelle Kahlenberg J. Cutaneous and Systemic Connections in Lupus. *Curr Opin Rheumatol* (2020) 32(6):583–9. doi: 10.1097/BOR.0000000000000739
17. Hochberg MC. Updating the American College of Rheumatology Revised Criteria for the Classification of Systemic Lupus Erythematosus. *Arthritis Rheum* (1997) 40(9):1725. doi: 10.1002/art.1780400928
18. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data. *Biostatistics* (2003) 4(2):249–64. doi: 10.1093/biostatistics/4.2.249
19. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The Sva Package for Removing Batch Effects and Other Unwanted Variation in High-Throughput Experiments. *Bioinformatics* (2012) 28(6):882–3. doi: 10.1093/bioinformatics/bts034
20. Law CW, Chen Y, Shi W, Smyth GK. Voom: Precision Weights Unlock Linear Model Analysis Tools for RNA-Seq Read Counts. *Genome Biol* (2014) 15(2): R29. doi: 10.1186/gb-2014-15-2-r29
21. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res* (2015) 43(7):e47. doi: 10.1093/nar/gkv007
22. Langfelder P, Horvath S. WGCNA: An R Package for Weighted Correlation Network Analysis. *BMC Bioinformatics* (2008) 9:559. doi: 10.1186/1471-2105-9-559
23. Aran D, Hu Z, Butte AJ. Xcell: Digitally Portraying the Tissue Cellular Heterogeneity Landscape. *Genome Biol* (2017) 18(1):220. doi: 10.1186/s13059-017-1349-1
24. Gudjonsson JE, Tsoi LC, Ma F, Billi AC, van Straalen KR, Vossen ARJV, et al. Contribution of Plasma Cells and B Cells to Hidradenitis Suppurativa Pathogenesis. *JCI Insight* (2020) 5(19). doi: 10.1172/jci.insight.139930
25. Kiefer K, Oropallo MA, Cancro MP, Marshak-Rothstein A. Role of Type I Interferons in the Activation of Autoreactive B Cells. *Immunol Cell Biol* (2012) 90(5):498–504. doi: 10.1038/icb.2012.10
26. Keller EJ, Patel NB, Patt M, Nguyen JK, Jørgensen TN. Partial Protection From Lupus-Like Disease by B-Cell Specific Type I Interferon Receptor Deficiency. *Front Immunol* (2020) 11:616064. doi: 10.3389/fimmu.2020.616064

27. Liu M, Guo Q, Wu C, Sterlin D, Goswami S, Zhang Y, et al. Type I Interferons Promote the Survival and Proinflammatory Properties of Transitional B Cells in Systemic Lupus Erythematosus Patients. *Cell Mol Immunol* (2019) 16 (4):367–79. doi: 10.1038/s41423-018-0010-6
28. Zhou S, Li Q, Zhao M, Lu L, Wu H, Lu Q. A Novel Humanized Cutaneous Lupus Erythematosus Mouse Model Mediated by IL-21-Induced Age-Associated B Cells. *J Autoimmun* (2021) 123:102686. doi: 10.1016/j.jaut.2021.102686
29. Mande P, Zirak B, Ko WC, Taravati K, Bride KL, Brodeur TY, et al. Fas Ligand Promotes an Inducible TLR-Dependent Model of Cutaneous Lupus-Like Inflammation. *J Clin Invest* (2018) 128(7):2966–78. doi: 10.1172/JCI98219
30. Magro CM, Segal JP, Crowson AN, Chadwick P. The Phenotypic Profile of Dermatomyositis and Lupus Erythematosus: A Comparative Analysis. *J Cutan Pathol* (2010) 37(6):659–71. doi: 10.1111/j.1600-0560.2009.01443.x
31. Quelhas da Costa R, Aguirre-Alastuey ME, Isenberg DA, Saracino AM. Assessment of Response to B-Cell Depletion Using Rituximab in Cutaneous Lupus Erythematosus. *JAMA Dermatol* (2018) 154(12):1432–40. doi: 10.1001/jamadermatol.2018.3793
32. Vital EM, Wittmann M, Edward S, Md Yusof MY, MacIver H, Pease CT, et al. Brief Report: Responses to Rituximab Suggest B Cell-Independent Inflammation in Cutaneous Systemic Lupus Erythematosus. *Arthritis Rheumatol* (2015) 67(6):1586–91. doi: 10.1002/art.39085

Conflict of Interest: JMK has received Grant support from Q32 Bio, Celgene/BMS, Ventus Therapeutics, and Janssen. JG has received Grant support from Celgene/BMS, Janssen, Eli Lilly, and Almirall. JMK has served on advisory boards for AstraZeneca, Eli Lilly, GlaxoSmithKline, Bristol Myers Squibb, Avion Pharmaceuticals, Provention Bio, Aurinia Pharmaceuticals, Ventus Therapeutics, and Boehringer Ingelheim. JG has served on advisory boards for AstraZeneca, Sanofi, Eli Lilly, Boehringer Ingelheim, Novartis, Janssen, Almirall, BMS.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers
in Immunology

ORIGINAL RESEARCH
published: 21 December 2021
doi: 10.3389/fimmu.2021.742080

Artificial Neural Network Analysis-Based Immune-Related Signatures of Primary Non-Response to Infliximab in Patients With Ulcerative Colitis

Xuanfu Chen[1], Lingjuan Jiang[2], Wei Han[3], Xiaoyin Bai[1], Gechong Ruan[1], Mingyue Guo[1], Runing Zhou[1], Haozheng Liang[1], Hong Yang[1] and Jiaming Qian[1]*

[1] Department of Gastroenterology, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China, [2] Medical Research Center, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China, [3] Department of Epidemiology and Biostatistics, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and School of Basic Medicine, Peking Union Medical College, Beijing, China

OPEN ACCESS

Edited by:
Xu-jie Zhou,
Peking University First Hospital, China

Reviewed by:
Carlo Chizzolini,
Université de Genève,
Switzerland
Zhixiu Li,
Queensland University of Technology,
Australia

***Correspondence:**
Hong Yang
Yangh@pumch.cn

Specialty section:
This article was submitted to Autoimmune and Autoinflammatory Disorders, a section of the journal Frontiers in Immunology

Received: *15 July 2021*
Accepted: *10 November 2021*
Published: *21 December 2021*

Citation:
Chen X, Jiang L, Han W, Bai X, Ruan G, Guo M, Zhou R, Liang H, Yang H and Qian J (2021) Artificial Neural Network Analysis-Based Immune-Related Signatures of Primary Non-Response to Infliximab in Patients With Ulcerative Colitis. Front. Immunol. 12:742080. doi: 10.3389/fimmu.2021.742080

Infliximab (IFX) is an effective medication for ulcerative colitis (UC) patients. However, one-third of UC patients show primary non-response (PNR) to IFX. Our study analyzed three Gene Expression Omnibus (GEO) datasets and used the RobustRankAggreg (RRA) algorithm to assist in identifying differentially expressed genes (DEGs) between IFX responders and non-responders. Then, an artificial intelligence (AI) technology, artificial neural network (ANN) analysis, was applied to validate the predictive value of the selected genes. The results showed that the combination of *CDX2*, *CHP2*, *HSD11B2*, *RANK*, *NOX4*, and *VDR* is a good predictor of patients' response to IFX therapy. The range of repeated overall area under the receiver-operating characteristic curve (AUC) was 0.850 ± 0.103. Moreover, we used an independent GEO dataset to further verify the value of the six DEGs in predicting PNR to IFX, which has a range of overall AUC of 0.759 ± 0.065. Since protein detection did not require fresh tissue and can avoid multiple biopsies, our study tried to discover whether the key information, analyzed by RNA levels, is suitable for protein detection. Therefore, immunohistochemistry (IHC) staining of colonic biopsy tissues from UC patients treated with IFX and a receiver-operating characteristic (ROC) analysis were used to further explore the clinical application value of the six DEGs at the protein level. The IHC staining of colon tissues from UC patients confirmed that VDR and RANK are significantly associated with IFX efficacy. Total IHC scores lower than 5 for VDR and lower than 7 for RANK had an AUC of 0.828 (95% CI: 0.665–0.991, p = 0.013) in predicting PNR to IFX. Collectively, we identified a predictive RNA model for PNR to IFX and explored an immune-related protein model based on the RNA model, including VDR and RANK, as a predictor of IFX non-response, and determined the cutoff value. The result showed a connection between the RNA and protein model, and both two models were available. However, the composite signature of VDR and RANK is more conducive to clinical application, which could be used to guide the preselection of patients who might benefit from pharmacological treatment in the future.

Keywords: infliximab, ulcerative colitis, artificial neural network analysis, prediction, primary non-response

INTRODUCTION

Ulcerative colitis (UC) is a chronic relapsing inflammatory disease of the colonic mucosa. UC is a relapsing disease requiring long-term management throughout life. The mainstay therapies for UC include 5-aminosalicylates, glucocorticoids, immunosuppressants, and biologic agents (1, 2). Biologic drugs, including antitumor necrosis factor (TNF)-α agents, anti-integrin drugs (vedolizumab), Janus kinase inhibitors (tofacitinib), and interleukin-12/23 antibodies (ustekinumab) (3), have driven UC therapy to a new era. The anti-TNF-α agent infliximab (IFX) is the oldest and most widely used biologic agent.

A meta-analysis showed that IFX was the highest-ranking biologic agent for the induction of clinical remission (OR 4.10, 95% CI: 2.58–6.52) and mucosal healing in moderate to severe UC (4, 5). However, according to previous studies, nearly one-third of UC patients show primary non-response (PNR). Moreover, studies have shown that other biologic agents have a higher failure rate in patients who previously failed to respond to IFX treatment than in those who are naïve to anti-TNF treatment (6, 7). Furthermore, the time PNR patients spend on IFX therapy can delay treatment, increase the risk of disease aggravation, and increase the economic burden of UC. Therefore, it is crucial to distinguish between PNR and effective responses to IFX treatment. Predictions of non-responses to IFX can assist the accurate selection of patients who could experience a clinical benefit and avoid potential adverse effects and unnecessary financial investment. Thus, an approach to identify markers from common, accessible samples, such as tissue biopsies or blood samples, is needed.

Previous studies have demonstrated that the therapeutic response depends on clinical factors, serum markers, and host genetics. Brandse et al. found that a high baseline serum level of C-reactive protein (CRP) was associated with lower serum concentrations of IFX, leading to non-response (8). Arias et al. identified a panel of serum markers (pANCA, CRP, and albumin) as independent predictors of the long-term outcome following IFX therapy in UC patients (9). Nevertheless, these indexes mainly related to disease activity and imperfectly predicted the primary therapeutic response to IFX (10). Burke et al. showed that genetic polymorphisms have predictive value for PNR to anti-TNF therapy in UC patients (11). Moreover, a high pretreatment expression of oncostatin M (OSM) was associated with anti-TNF resistance (12). However, the signatures of anti-TNF non-response mentioned above need further external clinical validation.

In the present study, we aimed to identify the specific markers underlying the PNR to IFX using combined datasets. Due to the expense of RNA sequencing, the RNA-seq dataset was small, and we used the bootstrapping method to randomly resample (13, 14). The first step to developing a predictor for clinical application is to find a repeatable result. We used an artificial intelligence (AI) technology, artificial neural network (ANN) analysis, to validate whether these data might be useful in estimating PNR. We repeated this process multiple times to validate the results. Moreover, we used another independent RNA dataset to confirm the result. Furthermore, an immunohistochemistry staining of colon tissue from UC patients who underwent IFX therapy was performed to explore the clinical application at the protein level. Ultimately, the findings of this work provide a greater understanding of which patients might receive therapeutic benefit from IFX therapy.

METHODS

Data Collection From the Gene Expression Omnibus Database

This study acquired clinical data and mRNA expression profiles of colon tissue from adult patients with UC from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). By using the keywords "ulcerative colitis" or "UC" and "IFX" or "infliximab," a total of eight series associated with UC treated by IFX were identified. After review, we selected three datasets (GSE12251, GSE16879, and GSE23597) containing the therapeutic efficacy of different dosages of IFX (15–18) as a discovery cohort. The platform used for the three datasets was the GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array. The selected patients all underwent colonoscopy, and biopsies of the diseased colon were performed before IFX therapy. Since the most commonly used dose of IFX in the clinic is 5 mg/kg and to maintain consistency among the three datasets, we selected patients who received a 5-mg/kg dose of IFX from the three datasets. Finally, 25 UC patients who responded to the first IFX treatment and 25 UC patients who exhibited PNR were included. The response was assessed in week 8 in the GSE12251 and GSE23597 datasets after the first infliximab treatment, and in weeks 4–6 in the GSE16879 dataset. The response definition was complete mucosal healing with a Mayo endoscopic subscore of 0 or 1 and a histological score of 0 or 1. An independent cohort from GEO (GSE73661) was used for further validation, which contained eight primary IFX responders and 15 non-responders. The response was assessed in weeks 4–6 in the GSE73661 dataset. The platform of the GSE73661 dataset was the GPL6244 [HuGene-1_0-st].

Data Extraction, Screening, and Aggregation of Differentially Expressed Genes

The pre-IFX-therapy-sequencing data of the obtained patients were extracted from the GSE12251, GSE16879, and GSE23597 datasets. UC patients who responded or did not respond to a 5-mg/kg dose of IFX at the first follow-up were selected and divided into the response group and PNR group. The limma R package (http://www.bioconductor.org/) was used to filter the Differentially Expressed Genes (DEGs) in each dataset. The same analysis was done in the validation cohort, the GSE73661 dataset. DEGs were defined as both an adjusted p-value < 0.05 and |log fold change (logFC)| > 0.5. The TXT files of all DEGs of the discovery datasets were sorted by logFC and saved for the subsequent integration analysis.

The three TXT files of all DEGs sorted by logFC were aggregated using the RobustRankAggreg (RRA) R package (https://CRAN.R-project.org/package=RobustRa-nkAggreg). The aggregated DEGs from all datasets, including upregulated and downregulated DEGs, were saved for subsequent analysis.

We selected aggregated upregulated and downregulated genes with a p-value lower than 0.05. Then, we ranked the genes by the logFC in order from the largest to the smallest. We reviewed the significant protein-coding DEGs and sorted out the genes expression in the alimentary tract through NCBI (https://www.ncbi.nlm.nih.gov/). We then reviewed published papers to determine genes which associate with immune activities to construct a list of proteins linked to the efficacy of IFX. Subsequently, we used the GSE16879 dataset, which contained sequencing data before and after IFX therapy, to determine the relationship between the selected protein-coding genes and IFX. Since the fewer indicators included, the higher the economic benefits obtained, we tried to find a better combination of DEGs.

Resampling Method and Artificial Neural Network Analysis

The subjects in the response group and PNR group were resampled by the “bootstrap” method. The dataset was randomly resampled to 250 by the proportion of the two groups (with replacement, i.e., when an item is sampled, it is immediately returned) (13). The samples from the resampling were analyzed by an ANN to show the efficiency of the model. To confirm the stability of the model, we repeated the resampling and ANN analysis 500 times. The process was also performed by shielding one input randomly. The range of area under the receiver-operating characteristic curve (AUC) was calculated. The same analysis parameters of ANN were used to verify the prediction ability of the selected DEGs in the validation dataset.

Exploring the Expression of the Selected DEGs at the Protein Level

Patients with UC receiving IFX monotherapy were enrolled from 2017 to 2020 at the Peking Union Medical College Hospital (PUMCH). Twenty-four UC patients were selected. The diagnostic criteria were based on the third European Crohn’s and Colitis Organization (ECCO) consensus guideline for UC and the 2018 Chinese consensus for inflammatory bowel disease (19, 20). We evaluated their clinical data at baseline, week 6, and week 14 after therapy. The response to IFX in week 6 was defined as a decrease in the partial Mayo score (Mayo score without endoscopy) of at least three points and at least 30% compared with the baseline data (21). A response in week 14 was defined as a decrease in the Mayo score of at least three points and at least 30% less than the baseline value, and the rectal bleeding score should decrease by more than 1 point or be equal to 0 or 1 point. The colonoscopic biopsies before the first IFX treatment of these patients were used for the immunohistochemistry (IHC) staining to verify the effectiveness of the obtained genes at the protein level. All colonic biopsy samples and clinical data of the patients used in this study were carried out with the approval of the Peking Union Medical College Hospital and the Chinese Academy Medical Science Ethics Committee (S-K1142).

Staining Off Target Proteins by Immunohistochemistry

We performed IHC staining off of the target proteins in formalin-fixed, paraffin-embedded colon tissues. The antigens were retrieved by boiling the samples for 10 min in 10 mM citrate (pH 6.0) or EDTA antigen repair solution (pH 9.0) (ZSGB-BIO). The slides were stained with rabbit monoclonal antibodies (Cell Signaling) and then incubated with a peroxidase-conjugated secondary antibody. Finally, the signals were visualized with diaminobenzidine (DAB) peroxidase substrate kit (Servicebio).

IHC Scoring

A flowchart of the IHC scoring and analysis is shown in **Figure 1**. The IHC staining was semiquantitatively evaluated by rating both the extent and intensity. First, we randomly selected 10 visual fields (×40) under a light microscope and counted 100 cells in each visual field. Then, we rated the extent as the proportion of positive cells on a scale from 0 to 4 as follows: 0, <1%; 1, 1% to 25%; 2, 26% to 50%; 3, 51% to 75%; and 4, ≥76%. Moreover, the intensity of the immunoreactivity (IR) was rated on a scale from 0 to 3 as follows: 0, no IR; 1, weak IR; 2, moderate IR; and 3, strong IR (**Table 1**) (22). We defined the IHC score of each protein as the mean value of the extent or intensity score in each visual field. The IHC scoring was analyzed independently by two gastroenterologists who were blinded to the patients’ response to IFX.

Additionally, we used a two-tailed exact Mann–Whitney U test (non-parametric) to compare the IHC score between the responders and non-responders. Any variable with a p-values lower than 0.05 in its extent or intensity scores was included in the multivariate analysis. Then, we defined the *total score* of each subject as the product of the mean extent and intensity score of each protein as follows:

$$\begin{aligned}&\textbf{Total IHC score of each protein in each sample}\\&= \textbf{mean extent score} \times \textbf{mean intensity score}\end{aligned}$$

Moreover, we used the bootstrap method and an ANN analysis to show the efficiency of the combination of the

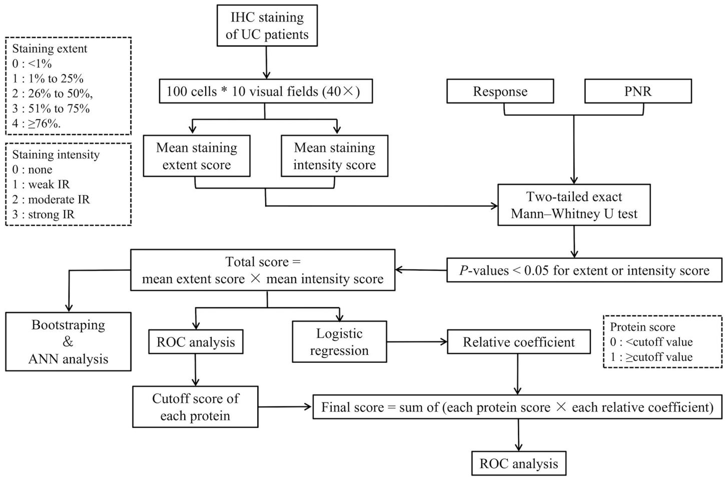


FIGURE 1 | Flowchart of the IHC scoring and analysis. IHC, immunohistochemistry; PNR, primary non-response; ANN analysis, artificial neural network analysis; ROC, receiver-operating characteristic.

selected proteins in predicting IFX efficacy. The resampling and ANN analysis process were repeated 500 times. The range of AUCs was calculated to measure the results of ANN analysis of the proteins in predicting the therapeutic effect of IFX in week 6 and week 14.

To achieve the threshold of distinction between a response and PNR, we divided the cutoff value of the total IHC score of each included protein by an ROC analysis. Moreover, the *protein score* was defined as 1 when the total IHC score was greater than or equal to the cutoff value and 0 when the total IHC score was lower than the cutoff value. Then, we used a logistic regression to calculate the relative coefficient of the IHC score of the proteins. We divided the regression coefficient of the other variables by the minimum regression coefficient and rounded the result to obtain the score of each variable. The product of the relative coefficient and protein score was obtained, and the sum of the products was defined as the final predictive score. An ROC curve was plotted to estimate the value of the selected proteins in predicting the therapeutic effect of IFX.

$$\begin{aligned}\textbf{Final score} &= \textbf{protein score 1} \times \textbf{relative coefficient 1}\\ &\quad + \textbf{protein score 2} \times \textbf{relative coefficient 2}\\ &\quad + \ldots + \textbf{protein score n} \times \textbf{relative coefficient n}\end{aligned}$$

TABLE 1 | Grading scale for the semiquantitative IHC scoring.

Score	Staining extent	Staining intensity
0	<1%	None
1	1%–25%	Weak immunoreactivity
2	26%–50%	Moderate immunoreactivity
3	51%–75%	Strong immunoreactivity
4	≥76%	–

Statistical Analysis

Non-parametric analyses were used to estimate the differences between the IFX response and non-response groups. The statistical tests were two-tailed and described in the figure legends. ROC curves were used to test the prediction value. All p-values less than 0.05 were considered significant. All analyses and the graph creation were performed in SPSS (version 25.0, IBM Corporation, Chicago, USA), R software (version 3.5.2, R Foundation for Statistical Computing, Vienna, Austria), and MATLAB (R2019a, MathWorks, USA).

RESULTS

Identification of DEGs Between Responders and Primary Non-Responders

According to the inclusion criteria for the sequencing data before 5 mg/kg IFX therapy, we extracted UC patients who were primary IFX responders or non-responders from the GSE12251, GSE16879, and GSE23597 datasets. The GSE12251 dataset included four responders and seven non-responders, the GSE16879 dataset contained eight responders and 16 non-responders, and the GSE23597 dataset included 13 responders and two non-responders (**Table 2**). The DEGs were screened using the limma R package (adjusted p-value < 0.05 and |logFC| > 0.5). The GSE12251 dataset contained 2,335 DEGs, including 1,346 upregulated genes and 989 downregulated genes. Furthermore, 934 upregulated genes and 852 downregulated genes were included in the GSE16879 dataset, resulting in a total of 1,786 DEGs in this dataset. Finally, the GSE23597 dataset contained 3,497 DEGs, including 1,390 upregulated genes and 2,107 downregulated genes. The DEGs in the three datasets are shown in **Table 3** and **Figure 2**.

TABLE 2 | Details of the UC patients receiving IFX therapy in the GEO database.

GEO dataset	Platform	PubMed ID	Sample	Time of biopsy	Time of assessment	Response	PNR
GSE12251	GPL570	19700435	Colonic tissue	Within 2 weeks before treatment	Week 8	4	7
GSE16879	GPL570	19956723	Colonic tissue	Within 1 week before treatment	Weeks 4–6	8	16
GSE23597	GPL570	21448149, 31039157	Colonic tissue	Within 2 weeks before treatment	Week 8	13	2

PNR, primary non-response.

Integrated DEGs Between Responders and Primary Non-Responders

The aggregated DEGs were screened by the RRA package (p-value < 0.05, $|\text{logFC}| > 0.5$). This method was based on the RRA algorithm in which each gene in each dataset was randomly arranged. If a gene ranked higher in all datasets, the associated p-value was lower, indicating that the possibility of this gene being a DEG was greater in all datasets. Using the RRA method, 624 integrated DEGs were identified, consisting of 18 upregulated genes and 606 downregulated genes. We selected the aggregated upregulated and downregulated DEGs by an associated p-value lower than 0.05, ranked the logFC in order from the largest to the smallest, identified the protein-coding genes, and determined the gene expression in the gastrointestinal tract by NCBI. Among the upregulated genes, those with low expression in the normal gastrointestinal tract were selected, while among the downregulated genes, those with high expression in the normal gastrointestinal tract were selected. We then reviewed published papers to consider proteins linked to immune or inflammatory processes. Ultimately, five downregulated proteins associated with PNR, including CDX2, CHP2, HSD11B2, RANK, and VDR, were selected; one upregulated protein, NOX4, was chosen. Furthermore, we used the GSE16879 dataset, which contained RNA sequencing data both before and after IFX treatment, to determine the relationship between the selected DEGs and IFX therapy. We found that 1) the non-responders to IFX tended to have a lower pretreatment expression of the downregulated DEGs compared with the responders; 2) the posttreatment expression of the downregulated DEGs displayed a trend of increases in the responders; and 3) the expression of the downregulated DEGs after treatment in those who responded to IFX was higher than that in those who did not respond to IFX. This phenomenon was the opposite in the upregulated DEG NOX4 (**Figure 3**). Thus, the following six proteins were ultimately selected for the construction of the predictive model of IFX efficacy: CDX2, CHP2, HSD11B2, RANK, NOX4, and VDR (**Figure 4**). CDX2, CHP2, HSD11B2, RANK, and VDR showed decreased expression in the non-responders, while NOX4 showed increased expression.

TABLE 3 | DEGs between the responders and non-responders in each dataset.

	Upregulated genes (p-value < 0.05 and logFC > 0.5)	Downregulated genes (p-value < 0.05 and logFC < -0.5)
GSE12251	1346	989
GSE16879	934	852
GSE23597	1390	2107

Resampling and ANN Analysis Results of the DEGs in the Discovery and Validation Cohort

We used the bootstrap method to randomly resample the response group (n = 25) and PNR group (n = 25) and enlarge the sample size to 250 in proportion of the two groups. Bootstrapping can reduce heterogeneity in different sample populations and avoid the problem of sample reduction caused by cross validation. Then, we used the resampled dataset to perform an ANN analysis (23, 24). The ANN analysis weighed the importance of the selected proteins, thus predicting the effect on achieving response to IFX therapy. Based on the collection of connected units, ANN loosely mimics neurons in the real brain. Each connection works as synapses in a biological brain. ANN can convey signals from one artificial neuron to another. Then, artificial neurons that receive signals can transmit these signals and signal additional artificial neurons connected to them. In typical ANN applications, the signals at a connection between artificial neurons are actual numbers and the outputs of each artificial neuron are calculated by a non-linear function of the sum of its inputs. Artificial neurons and their connections have a weight that adjusts as learning proceeds. The weight enhances or reduces the power of the signals at a connection in the ANN. ANN incorporates a system of interconnections based on simple mathematical models associated with learning algorithms. ANN consists of a four-layer (one input layer, two hidden layers, and one output layer) feedforward analysis. To develop the ANN, cases were randomly assigned to a training set (70%), test set (15%), and verification set (15%) through a generator of random numbers in our study. Backpropagation of error was applied as a learning rule by the online training method. The synaptic weights were calculated after each training data record.

As the more included the indicators, the higher economic burden for application, we tried different combinations of DEGs to find a better small protein combination. We performed the resampling and ANN analysis 500 times by selecting all integrated DEGs, top 300, top 100, top 50, and the six selected DEGs. The process was also performed by shielding one input randomly based on the six selected DEGs. The range of the repeated overall AUC of the six selected DEGs was 0.850 ± 0.103, which was similar to the different combination of the top DEGs (**Figure 5A** and **Supplementary Table 1**) and was slightly higher than that of the shielding-one-DEG model based on the six selected DEGs (**Figure 5B** and **Table 4**). The results showed that the six-DEG model had good economic benefits and performed better in predicting the IFX response. The repeated results demonstrated that the model was stable. We also performed an ANN analysis in the independent GEO dataset (GSE73661).

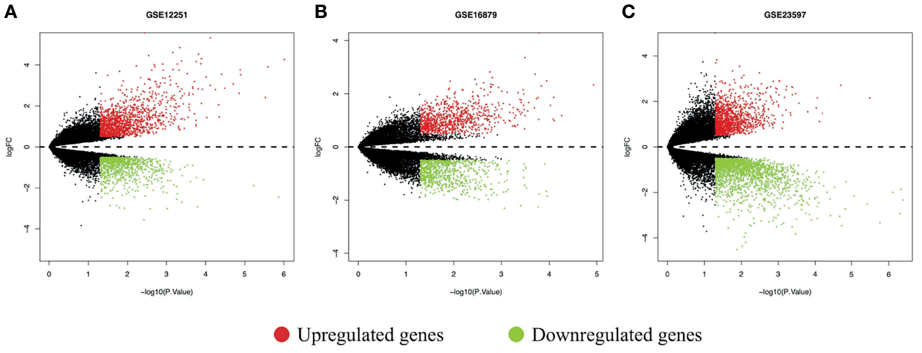


FIGURE 2 | DEGs between responders and non-responders in each dataset shown in volcano plots. **(A)** Volcano plot of the GSE12251 dataset; **(B)** volcano plot of the GSE16879 dataset; and **(C)** volcano plot of the GSE23597 dataset. The red dots represented upregulated genes based on a *p*-value < 0.05 and logFC > 0.5; the green dots represented downregulated genes based on a *p*-value < 0.05 and logFC<-0.5; the black spots represented genes with no significant difference in expression. DEGs, differentially expressed genes; logFC, log-fold change.

The results showed that the range of repeated overall AUC was 0.759 ± 0.065, indicating that the model was feasible.

Exploring Results of IHC in UC Patients Undergoing IFX Therapy

Biopsies are usually taken for pathological examination when UC patients undergo colonoscopy in the clinic. IHC analysis of clinical residual paraffin sections can avoid multiple biopsies and reduce the examination cost and time of patients. Thus, we tried to discover whether the key information, analyzed by RNA levels, is suitable for protein level detection. We used IHC analysis to explore the protein expression based on the selected DEGs and find clinical application predictors. Twenty-four UC patients were recruited from 2017 to 2020 at the Peking Union Medical College Hospital. Among these patients, 70.8% (n = 17) clinically responded to IFX treatment by week 6, and 29.2% (n = 7) did not. In addition, 54.17% (n = 13) of the patients achieved therapeutic benefits by week 14, while 45.83% (n = 11) did not. The proteins predicting IFX efficacy were evaluated by IHC scoring (**Figure 6**) without knowledge of the clinical data.

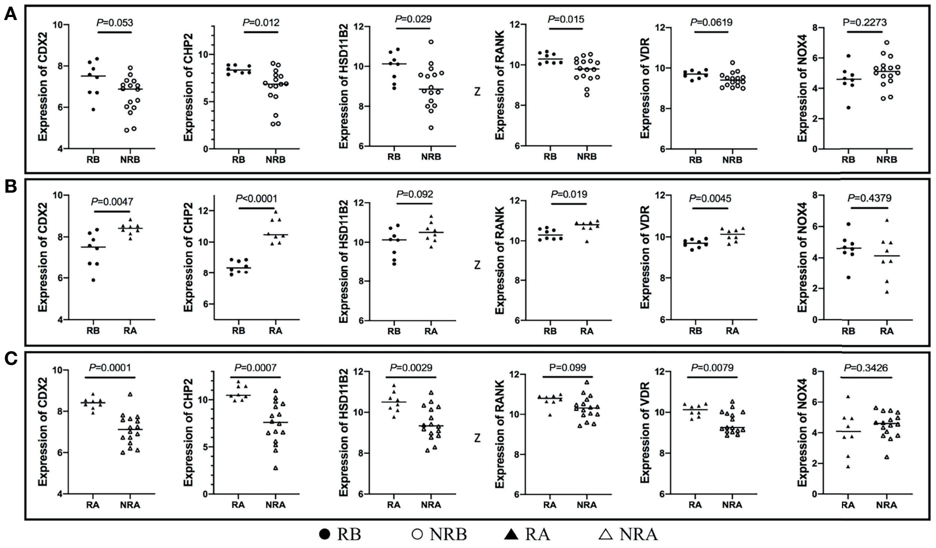


FIGURE 3 | DEG expression in different stages of IFX treatment. **(A)** The non-responders of IFX tended to have a lower pretreatment expression of the downregulated DEGs compared with the responders, and NOX4 displayed the opposite results; **(B)** the posttreatment expression of the downregulated DEGs exhibited a trend of increases in the responders, and NOX4 exhibited the opposite results; **(C)** the expression of the downregulated DEGs after treatment in those who responded to IFX was higher than that those who did not respond to IFX, and NOX4 exhibited the opposite results. DEGs, differentially expressed genes; IFX, infliximab; RB, sequencing data of responders before IFX therapy; NRB, sequencing data of non-responders before IFX therapy; RA, sequencing data of responders after IFX therapy; NRA, sequencing data of non-responders after IFX therapy.

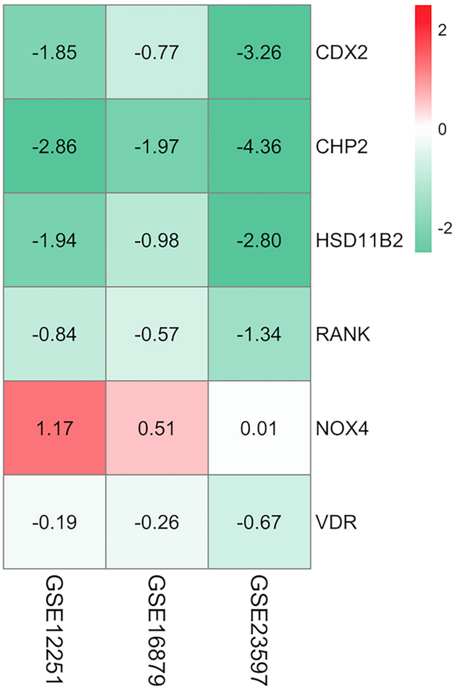


FIGURE 4 | Heatmap of the selected proteins. CDX2, CHP2, HSD11B2, RANK, and VDR displayed decreased expression, while NOX4 displayed increased expression; the red color represented logFC > 0, the green color represented logFC < 0 and the value in the box represented the logFC value. logFC, log fold change.

After the analysis, CHP2, HSD11B2, RANK, and VDR were found to have reduced mean IHC extent and intensity scores in the non-response group, and NOX4 had increased scores, which is consistent with the results of the analysis of the GEO datasets, while CDX2 had a limited difference between the groups. VDR and RANK statistically significantly differed between the two groups in terms of the intensity scores (*p*-value <0.05), and VDR showed a trend-level difference in terms of the extent scores (*p*-value = 0.065) (**Tables 5**, **6**). These two proteins were selected for further analysis.

We used the bootstrap method and an ANN analysis of VDR and RANK and repeated the analysis process 500 times. The AUC performed well in predicting the effect of IFX therapy. The range of repeated overall AUC was 0.837 ± 0.152 in predicting IFX efficacy in week 6 and was 0.776 ± 0.162 in predicting IFX efficacy in week 14 (**Figure 7** and **Supplementary Table 2**).

To determine the cutoff values for VDR and RANK, we used an ROC analysis. Ultimately, the cutoff value of the total IHC score was 5 for VDR and 7 for RANK. In addition, the logistic regression analysis showed that the regressive equation was as follows:

$$\mathbf{logit}\,(\mathbf{P}) = -\mathbf{0.799}\,(\textbf{total IHC score of VDR}) - \mathbf{0.44}\,(\textbf{total IHC score of RANK}) + \mathbf{5.024}$$

Therefore, the relative coefficient of VDR was 2, and that of RANK was 1. The final score of each sample was two times the protein score of VDR plus the protein score of RANK. The ROC curve was plotted to estimate the predictive value of the final score for IFX efficacy. The results showed that the final score had an IFX effective prediction value of 0.828 (95% CI: 0.665–0.991, *p*-value = 0.013) in week 6 (**Figure 8A**), with a sensitivity of 82.4% and a specificity of 71.4%. This finding indicates that total IHC scores less than 5 for VDR and less than 7 for RANK have good predictive value for primary non-response to IFX in patients with UC. The AUC was 0.759 (95% CI: 0.565–0.953, *p*-value = 0.032) in week 14 (**Figure 8B**), with a sensitivity of 69.2% and a specificity of 72.7%.

DISCUSSION

Precision medicine is becoming a hot topic in the medical literature in general, with oncology studies leading the way (25, 26). The most common strategy underlying all precision medicine

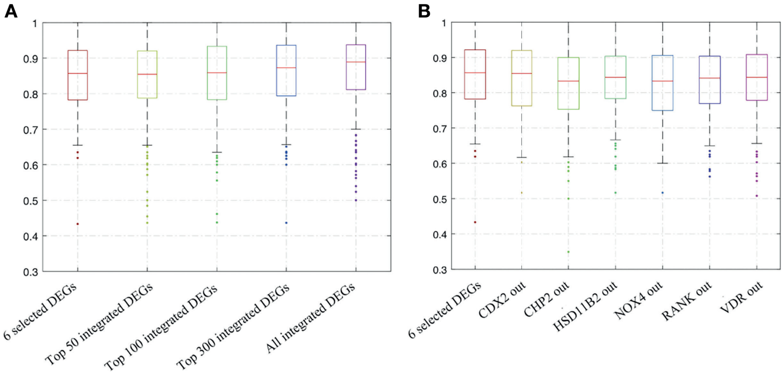


FIGURE 5 | Bootstrapping and ANN analysis results of the top DEGs, the six selected DEGs, and shielding of one DEG randomly based on the latter. **(A)** Analysis results of all integrated DEGs, top 300, top 100, top 50, and the six selected DEGs; **(B)** analysis results of shielding of one input randomly based on the six selected DEGs. ANN analysis, artificial neural network analysis; DEGs, differentially expressed genes.

TABLE 4 | AUC of different combinations of the six selected DEGs.

DEGs combination	AUC (mean ± SD)
Six selected DEGs	0.850 ± 0.103
CDX2_out	0.837 ± 0.106
CHP2_out	0.823 ± 0.115
HSD11B2_out	0.833 ± 0.100
NOX4_out	0.829 ± 0.105
RANK_out	0.836 ± 0.100
VDR_out	0.838 ± 0.103

is that distinct patient characteristics are used to tailor the therapeutic tactics, with the help of biomarker profiles (27). Our study extracted DEGs from a publicly available database and identified several gene signatures of patients diagnosed with UC with primary non-response to IFX based on the RRA algorithm, gastrointestinal expression, and previous studies, including *CDX2*, *CHP2*, *HSD11B2*, *RANK*, *NOX4*, and *VDR*. We used the bootstrap method and an ANN analysis to confirm that the markers were repeatable for clinical application. Moreover, an independent GEO cohort was used to verify the result. We also used samples from UC patients to explore the protein expression based on the selected DEGs. The result showed a connection between the RNA and protein model, and both two models were available, but the protein model is more reliable and more conducive to clinical application. Finally, total IHC scores less than 5 for VDR and less than 7 for RANK jointly achieved an AUC of 0.828 (95% CI: 0.665–0.991, *p*-value = 0.013) in predicting PNR to IFX. The ANN analysis further confirmed these results.

UC is a chronic inflammatory disease with an increasing incidence worldwide, affecting more than 1 million individuals in Western countries and many more globally (1, 28). UC carries a life-long risk of morbidity, especially in the moderate-to-severe disease stage. Thus far, an increasing number of biologics agents have been used for UC treatment in the clinic, including IFX, vedolizumab, adalimumab, and ustekinumab. The application of biological agents benefits patients in many aspects (3, 29). Previous studies have shown that biological agents are more effective than traditional medications in terms of short-term response (OR = 4.01, 95% CI 3.08–5.23), long-term remission (OR = 2.80, 95% CI = 1.89–4.14), severe UC rescue, and colectomy rate reduction (29.2% versus 58.3%; *p* = 0.017) (21, 30–32). A meta-analysis showed that IFX was the most effective agent at inducing remission in biologic-naive patients with moderate to severe UC (33).

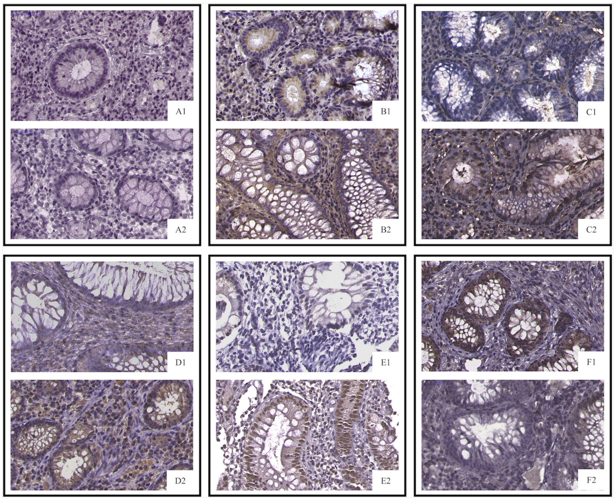


FIGURE 6 | IHC staining of selected proteins. (**A1**, PNR; **A2**, response) CDX2 did not differ between the primary IFX non-responders and responders; (**B1**, PNR; **B2,** response) CHP2 (**C1**, PNR; **C2**, response), HSD11B2 (**D1**, PNR; **D2**, response), RANK (**E1**, PNR; **E2**, response), and VDR staining was decreased in the primary non-responders, while NOX4 (**F1**, PNR; **F2**, response) was increased in the non-responders. PNR, primary non-response.

Nevertheless, treatment resistance remains a tremendous clinical challenge for UC patients. As the most cost-effective biologic (34), IFX shows significant curative efficacy, but close to one-third of UC patients are primary non-responders to this drug. Moreover, prior exposure to IFX may decrease the efficacy of other biologics (6, 7, 35). As IFX is most widely used in patients with moderate to severe UC, the failure of this drug as a first-line therapy could delay the onset of effective treatment. Therefore, personalized therapy for UC and predictive methods of individual response to IFX therapy are urgently needed (10). Our research responds to this pressing need and is expected to yield practical benefits in precision medicine for UC.

Six protein-coding genes predicting IFX efficacy were initially included in our study. Mostly those in the previous studies are clinical indicators, which predict IFX efficacy by responding to disease activity of UC (8, 9). Our study focuses more on predicting primary unresponsiveness than other clinical indicators and might reveal the mechanism of IFX therapeutic effects from the molecular level or pathway. Since protein expression is not always correlated with mRNA expression and protein level detection does not require fresh tissue and can avoid multiple biopsies, we used IHC to further explore the protein expression results in another dataset.

The protein-coding genes involved are strongly correlated with changes in the immune-based response and different immune cell types, including macrophages, dendritic cells (DCs), and T cells. CDX2, a transcription factor, has been shown to have a decreased expression in UC (36), play an essential role in intestinal homeostasis, and act as a context-dependent tumor suppressor in colorectal cancer. The deletion of CDX2 from the intestinal epithelium in mice leads to macrophage infiltration, causing chronic inflammatory responses (37). However, CDX2 did not revert to normal in CD patients treated with anti-TNF-α biologics (38). In our study, CDX2 did not differ between the groups by IHC. The biological function of CHP2 remains largely unknown. Guo-Dong Li et al. found that CHP2 can increase the nuclear presence of nuclear factor of activated T cells (NFATc3) and enhance activated T cell activity (39). In particular, T helper (Th) 2-mediated inflammation plays a role in UC (40). NFATs can cooperate with various transcription factors to form transcriptional complexes and integrate signaling pathways to change transcriptional patterns (41, 42). HSD11B2 and NOX4 are enriched in the hypoxia response. Tissue hypoxia, which decreases HSD11B2 and increases NOX4 expression, occurs in

TABLE 5 | The extent of the staining of IFX efficacy-predicting proteins in colonic biopsies from UC patients.

Variable	Staining extent score					Mean score	*p*-value[a]
	0 (n, %)	1 (n, %)	2 (n, %)	3 (n, %)	4 (n, %)		
CDX2							0.757
Responders	13 (76.5%)	4 (23.5%)	0	0	0	0.24	
Non-responders	5 (71.4%)	1 (14.3%)	1 (14.3%)	0	0	0.34	
HSD11B2							0.534
Responders	1 (5.9%)	6 (35.3%)	5 (29.4%)	4 (23.5%)	1 (5.9%)	1.88	
Non-responders	0	4 (57.1%)	2 (28.6%)	1 (14.3%)	0	1.57	
CHP2							0.209
Responders	0	3 (17.6%)	4 (23.5%)	5 (29.4%)	5 (29.4%)	2.71	
Non-responders	0	0	6 (85.7%)	1 (14.3%)	0	2.14	
RANK							0.114
Responders	0	0	6 (35.3%)	5 (29.4%)	6 (35.3%)	3.00	
Non-responders	0	1 (14.3%)	3 (42.9%)	3 (42.9%)	0	2.29	
NOX4							0.234
Responders	0	0	12 (70.6%)	5 (29.4%)	0	2.29	
Non-responders	0	0	3 (42.9%)	3 (42.9%)	1 (14.3%)	2.71	
VDR							0.065
Responders	0	5 (29.4%)	8 (47.1%)	3 (17.6%)	1 (5.9%)	2.00	
Non-responders	0	5 (71.4%)	2 (28.6%)	0	0	1.29	

[a] *A Mann–Whitney U test was used for the analysis.*

chronic inflammatory conditions, such as IBD. Van Welden et al. suggested that hypoxia of the colonic mucosa activates hypoxia inducible factors (HIFs) and the regulation of nuclear factor κB (NF-κB) (43). Yu et al. found that HIF-1α was upregulated in UC patients and positively related to disease progression (44). Therefore, colonic tissue hypoxia and hypoxia-induced signaling may be detection and therapeutic targets in UC (43). The reduction in HSD11B2 and the increase in NOX4 suggest a higher hypoxia response, which regulates inflammatory and immune processes and results in a complex hypoxia-immune-based microenvironment. Despite the expression of CDX2, CHP2, HSD11B2, and NOX4 related to IFX therapy and coping with inflammatory activity, their protein expression did not show a difference in the validation cohort. This finding might account for the different disease complexities and activities of UC patients between the public datasets and our enrolled subjects. We did not include these proteins in the protein prediction model.

Regarding the ultimately involved proteins, several reports from our group and others have highlighted the importance of VDR, a receptor of vitamin D, in UC. The colonic expression of VDR was inversely associated with disease activity in UC (45). Moreover, in our previous research, 25[OH]D3 levels were negatively correlated with the disease severity of UC ($r = -0.371$, $p < 0.001$) (46). A study by Shirwaikar Thomas et al. showed that

TABLE 6 | The intensity of the staining of IFX efficacy-predicting proteins in colonic biopsies from UC patients.

Variable	Staining intensity score				Mean score	*p*-value[a]
	0 (n, %)	1 (n, %)	2 (n, %)	3 (n, %)		
CDX2						0.757
Responders	13 (76.5%)	4 (23.5%)	0	0	0.24	
Non-responders	5 (71.4%)	1 (14.3%)	1 (14.3%)	0	0.43	
HSD11B2						0.166
Responders	1 (5.9%)	5 (29.4%)	8 (47.1%)	3 (17.6%)	1.76	
Non-responders	0	5 (71.4%)	2 (28.6%)	0	1.29	
CHP2						0.455
Responders	0	0	10 (58.8%)	7 (41.2%)	2.41	
Non-responders	0	1 (14.3%)	4 (57.1%)	2 (28.6%)	2.14	
RANK						0.034
Responders	0	0	7 (41.2%)	10 (58.8%)	2.59	
Non-responders	0	2 (28.6%)	4 (57.1%)	1 (14.3%)	1.86	
NOX4						0.349
Responders	0	1 (5.9%)	8 (47.1%)	8 (47.1%)	2.41	
Non-responders	0	0	2 (28.6%)	5 (71.4%)	2.71	
VDR						0.024
Responders	0	0	2 (11.8%)	15 (88.2%)	2.88	
Non-responders	0	0	5 (71.4%)	2 (28.6%)	2.29	

[a] *A Mann–Whitney U test was used for the analysis.*

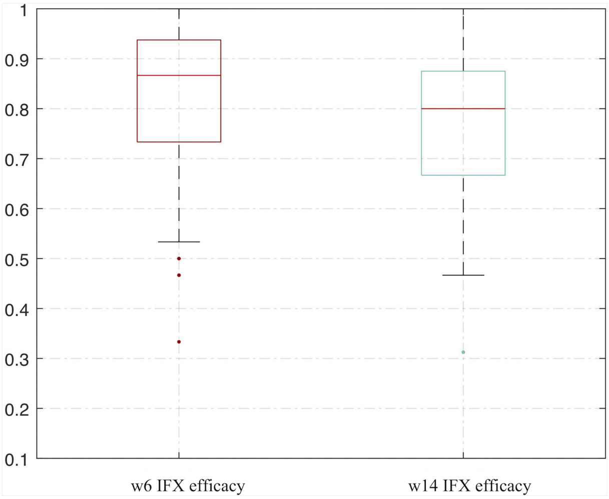


FIGURE 7 | Bootstrapping and ANN analysis results of VDR and RANK in predicting IFX efficacy in week 6 and week 14. ANN analysis, artificial neural network analysis; IFX, infliximab.

in IBD patients, those with active endoscopic inflammation have a lower vitamin D level than those in remission (47). Furthermore, low pretreatment serum 25[OH]D predicted vedolizumab failure in patients with IBD (48).

UC results from T helper (Th) 2-mediated inflammation, leading to the possibility that inhibitors of Th2 cytokines might be helpful in the treatment of UC (49, 50). Vitamin D has been shown to inhibit the proliferation of T cells from patients with active UC (51), which might reduce Th2 cell-induced inflammation. Furthermore, the levels of Th2 cells were higher in anti-TNF-non-responders in UC (52). A study by Song et al. demonstrated that VDR restricts Th2-biased inflammation in the heart (53). Therefore, the reduction in VDR in colonic tissue might correlate with a strengthening of Th2-mediated inflammation and anti-TNF non-response. Bingning et al. showed that VDR activation performs a solid anti-inflammatory function in macrophages and ameliorates insulin resistance (54). VDR signaling in macrophages suppresses NF-κB activity and reduces inflammatory factor interactions (55). VDR also regulates the function of Paneth cells in releasing antimicrobial peptides to modulate the innate immune process. Thus, the regulation of VDR on immune cells might improve intestinal inflammation, leading to disease activity.

Receptor activator of nuclear factor κB (RANK), also known as TNFRSF11A, is a member of the TNF receptor superfamily. The interactions between RANK and its ligand (RANKL) regulate T cell/DC communications, DC survival, and naive T cell proliferation (56, 57). Previous studies have shown that UC is characterized by an increase in activated T cells and T-regulatory cells and a decrease in naive T-cells (58, 59). DCs monitor the surrounding microenvironment, sample antigens, and induce tolerance or incite a host defense proinflammatory response in UC (60). Therefore, a reduction in RANK might lead to an imbalance in the immune microenvironment by affecting DCs and T cells, thereby inducing UC activity.

Collectively, our study demonstrates that total IHC scores less than 5 for VDR and less than 7 for RANK were associated with non-response to IFX. The diminished expression of VDR and RANK may account for the immune-related changes in the intestinal microenvironment and reduce anti-inflammatory factors, leading to an increase in disease activity. Meanwhile, the modulation of different immune cell populations and inflammatory processes may lower anti-inflammatory cell types and weaken the immune response. Therefore, IFX may not be sufficiently robust to address this complicated inflammatory status, resulting in an inadequate therapeutic effect.

Our study has several strengths. First, we obtained transcriptome data from public datasets for the integration analysis, which is the premise of precision medicine. Second, the resampling method was used to expand the data, and then an ANN analysis was used for internal verification and prediction. We

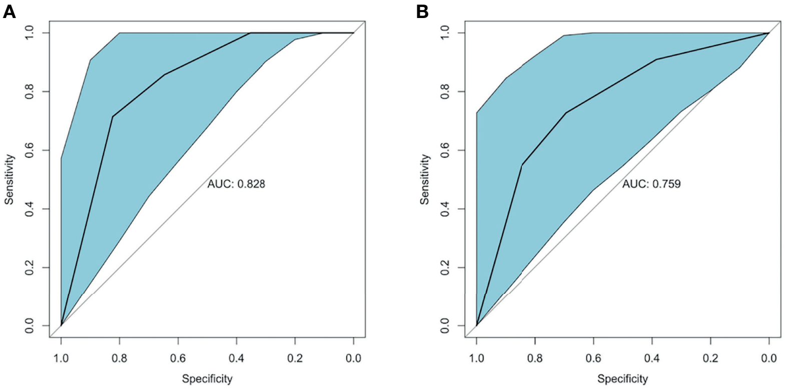


FIGURE 8 | ROC curve of VDR and RANK in predicting IFX efficacy in week 6 and week 14. **(A)** The AUC of the estimation 6 weeks after IFX treatment was 0.828. **(B)** The AUC of the estimation 14 weeks after IFX treatment was 0.759. ROC, receiver-operating characteristic; AUC, area under the receiver-operating characteristic curve; IFX, infliximab.

repeated the analysis process many times to show the stability of the model, forming a foundation for clinical application in the prediction of PNR. The significant proteins are readily tested in practice and are convenient for clinical application. In addition, we verified the validity of the protein profiles by IHC staining of colonic tissues from UC patients treated with IFX in our hospital. In previous studies, clinical factors, serum markers, and host genetics were demonstrated to play a role in the therapeutic response but did not accurately predict PNR. The secondary validation process in our study demonstrated that the clinical application of the immune-related signatures of primary IFX non-response in UC patients is repeatable. Furthermore, the time point of the response assessment was 6 to 8 weeks after the first IFX treatment in the GEO datasets and our enrolled subjects. A previous study showed that early measurement could better predict future remission and, thus, possibly benefit decision making (61).

Our research is not without limitations. To maintain consistency with the GEO databases, the clinical Mayo score 6–8 weeks after IFX treatment was used as the assessment when the recruited UC patients did not have endoscopy data. Thus, our study showed evidence of consistency and presented early predictive value even when an endoscopic evaluation was unavailable. However, our method may miss some patients whose endoscopic response is better or worse than their clinical response, which could increase the false-positive rate or the false-negative rate of the external verification. To reduce this bias, we also estimated the therapeutic efficacy in week 14 (**Figure 5B**), which included an endoscopic score. The signatures also showed good predictive value, with an AUC of 0.759. Although we identified the thresholds for VDR and RANK in predicting IFX efficacy, the results showed minor differences and overlap to some extent to distinguish responders and non-responders. However, our study provides preliminary data for using proteins to predict IFX efficacy. In the future, other more sensitive protein identification methods, such as electrical detection methodologies, might be developed for the precision treatment in the clinical practice (62). Furthermore, the percentage of non-responding patients in week 14 was higher than that in week 6, indicating that early assessment is preferable as an aid for decision making. Nevertheless, large-scale prospective studies are needed to correct this limitation.

In conclusion, this study found that total IHC scores less than 5 for VDR and less than 7 for RANK were good immune-based protein signatures of PNR to anti-TNF treatment in UC patients. Applying this panel in clinical practice could help clinicians identify likely IFX non-responders before initiating therapy. Nevertheless, the practical advantage of such a tailored approach needs to be confirmed in the future.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Peking Union Medical College Hospital and the Chinese Academy Medical Science Ethics Committee (S-K1142). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

XC and HY conceptualized and designed the research. XC carried out the data analysis. XC and RZ carried out the experimental procedures. HY and JQ oversaw the study and provided financial support. This manuscript was reviewed and revised by LJ, WH, XB, GR, MG, and HL. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the National Natural Science Foundation of China (No. 81570505 and No. 81970495), the Natural Science Foundation of Beijing Municipality (No. 7202161), the Health Research & Special Projects Grant of China (No. 201002020 and No. 201502005), the Peking Union Medical College Postgraduate Innovation Fund (No. 2019-1002-71), and the National Clinical Specialty Projects along with the CAMS Innovation Fund for Medical Sciences (No. 2021-1-I2M-001).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.742080/full#supplementary-material

REFERENCES

1. Rubin DT, Ananthakrishnan AN, Siegel CA, Sauer BG, Long MD. ACG Clinical Guideline: Ulcerative Colitis in Adults. *Am J Gastroenterol* (2019) 114 (3):384–413. doi: 10.14309/ajg.0000000000000152
2. Feuerstein JD, Isaacs KL, Schneider Y, Siddique SM, Falck-Ytter Y, Singh S, et al. AGA Clinical Practice Guidelines on the Management of Moderate to Severe Ulcerative Colitis. *Gastroenterology* (2020) 158(5):1450–61. doi: 10.1053/j.gastro.2020.01.006
3. Sands BE, Sandborn WJ, Panaccione R, O'Brien CD, Zhang H, Johanns J, et al. Ustekinumab as Induction and Maintenance Therapy for Ulcerative Colitis. *N Engl J Med* (2019) 381(13):1201–14. doi: 10.1056/NEJMoa1900750
4. Singh S, Fumery M, Sandborn WJ, Murad MH. Systematic Review With Network Meta-Analysis: First- and Second-Line Pharmacotherapy for Moderate-Severe Ulcerative Colitis. *Aliment Pharmacol Ther* (2018) 47 (2):162–75. doi: 10.1111/apt.14422
5. Danese S, Fiorino G, Peyrin-Biroulet L, Lucenteforte E, Virgili G, Moja L, et al. Biological Agents for Moderately to Severely Active Ulcerative Colitis:

A Systematic Review and Network Meta-Analysis. *Ann Intern Med* (2014) 160(10):704–11. doi: 10.7326/M13-2403

6. Favale A, Onali S, Caprioli F, Pugliese D, Armuzzi A, Macaluso FS, et al. Comparative Efficacy of Vedolizumab and Adalimumab in Ulcerative Colitis Patients Previously Treated With Infliximab. *Inflamm Bowel Dis* (2019) 25(11):1805–12. doi: 10.1093/ibd/izz057
7. Kopylov U, Verstockt B, Biedermann L, Sebastian S, Pugliese D, Sonnenberg E, et al. Effectiveness and Safety of Vedolizumab in Anti-TNF-Naïve Patients With Inflammatory Bowel Disease-A Multicenter Retrospective European Study. *Inflamm Bowel Dis* (2018) 24(11):2442–51. doi: 10.1093/ibd/izy155
8. Brandse JF, Mathôt RA, van der Kleij D, Rispens T, Ashruf Y, Jansen JM, et al. Pharmacokinetic Features and Presence of Antidrug Antibodies Associate With Response to Infliximab Induction Therapy in Patients With Moderate to Severe Ulcerative Colitis. *Clin Gastroenterol Hepatol* (2016) 14(2):251–8.e1-2. doi: 10.1016/j.cgh.2015.10.029
9. Arias MT, Vande Casteele N, Vermeire S, de Buck van Overstraeten A, Billiet T, Baert F, et al. A Panel to Predict Long-Term Outcome of Infliximab Therapy for Patients With Ulcerative Colitis. *Clin Gastroenterol Hepatol* (2015) 13(3):531–8. doi: 10.1016/j.cgh.2014.07.055
10. Denson LA, Curran M, McGovern DPB, Koltun WA, Duerr RH, Kim SC, et al. Challenges in IBD Research: Precision Medicine. *Inflamm Bowel Dis* (2019) 25(Suppl 2):S31–9. doi: 10.1093/ibd/izz078
11. Burke KE, Khalili H, Garber JJ, Haritunians T, McGovern DPB, Xavier RJ, et al. Genetic Markers Predict Primary Nonresponse and Durable Response to Anti-Tumor Necrosis Factor Therapy in Ulcerative Colitis. *Inflamm Bowel Dis* (2018) 24(8):1840–8. doi: 10.1093/ibd/izy083
12. West NR, Hegazy AN, Owens BMJ, Bullers SJ, Linggi B, Buonocore S, et al. Oncostatin M Drives Intestinal Inflammation and Predicts Response to Tumor Necrosis Factor-Neutralizing Therapy in Patients With Inflammatory Bowel Disease. *Nat Med* (2017) 23(5):579–89. doi: 10.1038/nm.4307
13. Henderson AR. The Bootstrap: A Technique for Data-Driven Statistics. Using Computer-Intensive Analyses to Explore Experimental Data. *Clin Chim Acta* (2005) 359(1-2):1–26. doi: 10.1016/j.cccn.2005.04.002
14. Al Seesi S, Tiagueu YT, Zelikovsky A, Măndoiu II. Bootstrap-Based Differential Gene Expression Analysis for RNA-Seq Data With and Without Replicates. *BMC Genomics* (2014) 15(Suppl 8):S2. doi: 10.1186/1471-2164-15-S8-S2
15. Arijs I, Li K, Toedter G, Quintens R, Van Lommel L, Van Steen K, et al. Mucosal Gene Signatures to Predict Response to Infliximab in Patients With Ulcerative Colitis. *Gut* (2009) 58(12):1612–9. doi: 10.1136/gut.2009.178665
16. Arijs I, De Hertogh G, Lemaire K, Quintens R, Van Lommel L, Van Steen K, et al. Mucosal Gene Expression of Antimicrobial Peptides in Inflammatory Bowel Disease Before and After First Infliximab Treatment. *PloS One* (2009) 4(11):e7984. doi: 10.1371/journal.pone.0007984
17. Toedter G, Li K, Marano C, Ma K, Sague S, Huang CC, et al. Gene Expression Profiling and Response Signatures Associated With Differential Responses to Infliximab Treatment in Ulcerative Colitis. *Am J Gastroenterol* (2011) 106(7):1272–80. doi: 10.1038/ajg.2011.83
18. Pavlidis S, Monast C, Loza MJ, Branigan P, Chung KF, Adcock IM, et al. I_MDS: An Inflammatory Bowel Disease Molecular Activity Score to Classify Patients With Differing Disease-Driving Pathways and Therapeutic Response to Anti-TNF Treatment. *PloS Comput Biol* (2019) 15(4):e1006951. doi: 10.1371/journal.pcbi.1006951
19. Inflammatory Bowel Disease Group, Chinese Society of Gastroenterology and Chinese Medical Association. Chinese Consensus on Diagnosis and Treatment of Inflammatory Bowel Disease (Beijing, 2018). *J Dig Dis* (2021) 22(6):298–317. doi: 10.1111/1751-2980.12994
20. Magro F, Gionchetti P, Eliakim R, Ardizzone S, Armuzzi A, Barreiro-de Acosta M, et al. Third European Evidence-Based Consensus on Diagnosis and Management of Ulcerative Colitis. Part 1: Definitions, Diagnosis, Extra-Intestinal Manifestations, Pregnancy, Cancer Surveillance, Surgery, and Ileo-Anal Pouch Disorders. *J Crohns Colitis* (2017) 11(6):649–70. doi: 10.1093/ecco-jcc/jjx008
21. Moss AC, Farrell RJ. Infliximab for Induction and Maintenance Therapy for Ulcerative Colitis. *Gastroenterology* (2006) 131(5):1649–51. doi: 10.1053/j.gastro.2006.09.039
22. Conner JR, Hirsch MS, Jo VY. Hnf1β and S100A1 are Useful Biomarkers for Distinguishing Renal Oncocytoma and Chromophobe Renal Cell Carcinoma in FNA and Core Needle Biopsies. *Cancer Cytopathol* (2015) 123(5):298–305. doi: 10.1002/cncy.21530
23. LeCun Y, Bengio Y, Hinton G. Deep Learning. *Nature* (2015) 521(7553):436–44. doi: 10.1038/nature14539
24. Indini A, Di Guardo L, Cimminiello C, De Braud F, Del Vecchio M. Artificial Intelligence Estimates the Importance of Baseline Factors in Predicting Response to Anti-PD1 in Metastatic Melanoma. *Am J Clin Oncol* (2019) 42(8):643–8. doi: 10.1097/COC.0000000000000566
25. Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Curran WJ Jr, Wu YL, et al. Lung Cancer: Current Therapies and New Targeted Treatments. *Lancet* (2017) 389(10066):299–311. doi: 10.1016/S0140-6736(16)30958-8
26. Johnson TM. Perspective on Precision Medicine in Oncology. *Pharmacotherapy* (2017) 37(9):988–9. doi: 10.1002/phar.1975
27. König IR, Fuchs O, Hansen G, von Mutius E, Kopp MV. What is Precision Medicine? *Eur Respir J* (2017) 50(4):1700391. doi: 10.1183/13993003.00391-2017
28. Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, et al. Worldwide Incidence and Prevalence of Inflammatory Bowel Disease in the 21st Century: A Systematic Review of Population-Based Studies. *Lancet* (2017) 390(10114):2769–78. doi: 10.1016/S0140-6736(17)32448-0
29. Feagan BG, Rutgeerts P, Sands BE, Hanauer S, Colombel JF, Sandborn WJ, et al. Vedolizumab as Induction and Maintenance Therapy for Ulcerative Colitis. *N Engl J Med* (2013) 369(8):699–710. doi: 10.1056/NEJMoa1215734
30. Sandborn WJ, Rutgeerts P, Feagan BG, Reinisch W, Olson A, Johanns J, et al. Colectomy Rate Comparison After Treatment of Ulcerative Colitis With Placebo or Infliximab. *Gastroenterology* (2009) 137(4):1250–520. doi: 10.1053/j.gastro.2009.06.061
31. Guo C, Wu K, Liang X, Liang Y, Li R. Infliximab Clinically Treating Ulcerative Colitis: A Systematic Review and Meta-Analysis. *Pharmacol Res* (2019) 148:104455. doi: 10.1016/j.phrs.2019.104455
32. Järnerot G, Hertervig E, Friis-Liby I, Blomquist L, Karlén P, Grännö C, et al. Infliximab as Rescue Therapy in Severe to Moderately Severe Ulcerative Colitis: A Randomized, Placebo-Controlled Study. *Gastroenterology* (2005) 128(7):1805–11. doi: 10.1053/j.gastro.2005.03.003
33. Singh S, Murad MH, Fumery M, Dulai PS, Sandborn WJ. First- and Second-Line Pharmacotherapies for Patients With Moderate to Severely Active Ulcerative Colitis: An Updated Network Meta-Analysis. *Clin Gastroenterol Hepatol* (2020) 18(10):2179–2191.e6. doi: 10.1016/j.cgh.2020.01.008
34. Petryszyn P, Ekk-Cierniakowski P, Zurakowski G. Infliximab, Adalimumab, Golimumab, Vedolizumab and Tofacitinib in Moderate to Severe Ulcerative Colitis: Comparative Cost-Effectiveness Study in Poland. *Therap Adv Gastroenterol* (2020) 13:1756284820941179. doi: 10.1177/1756284820941179
35. Welty M, Mesana L, Padhiar A, Naessens D, Diels J, van Sanden S, et al. Efficacy of Ustekinumab vs. Advanced Therapies for the Treatment of Moderately to Severely Active Ulcerative Colitis: A Systematic Review and Network Meta-Analysis. *Curr Med Res Opin* (2020) 36(4):595–606. doi: 10.1080/03007995.2020.1716701
36. Coskun M. The Role of CDX2 in Inflammatory Bowel Disease. *Dan Med J* (2014) 61(3):B4820.
37. Chewchuk S, Jahan S, Lohnes D. Cdx2 Regulates Immune Cell Infiltration in the Intestine. *Sci Rep* (2021) 11(1):15841. doi: 10.1038/s41598-021-95412-w
38. Younes M, Rahimi E, DuPont AW, Ly CJ, Ertan A. Anti-TNF-α Biologics Do Not Reverse CDX2 Downregulation in Patients With Crohn's Disease. *Ann Clin Lab Sci* (2020) 50(2):172–4.
39. Li GD, Zhang X, Li R, Wang YD, Wang YL, Han KJ, et al. CHP2 Activates the Calcineurin/Nuclear Factor of Activated T Cells Signaling Pathway and Enhances the Oncogenic Potential of HEK293 Cells. *J Biol Chem* (2008) 283(47):32660–8. doi: 10.1074/jbc.M806684200
40. Geremia A, Biancheri P, Allan P, Corazza GR, Di Sabatino A. Innate and Adaptive Immunity in Inflammatory Bowel Disease. *Autoimmun Rev* (2014) 13(1):3–10. doi: 10.1016/j.autrev.2013.06.004
41. Wu Y, Borde M, Heissmeyer V, Feuerer M, Lapan AD, Stroud JC, et al. FOXP3 Controls Regulatory T Cell Function Through Cooperation With NFAT. *Cell* (2006) 126(2):375–87. doi: 10.1016/j.cell.2006.05.042
42. Smith-Garvin JE, Koretzky GA, Jordan MS. T Cell Activation. *Annu Rev Immunol* (2009) 27:591–619. doi: 10.1146/annurev.immunol.021908.132706
43. Van Welden S, Selfridge AC, Hindryckx P. Intestinal Hypoxia and Hypoxia-Induced Signalling as Therapeutic Targets for IBD. *Nat Rev Gastroenterol Hepatol* (2017) 14(10):596–611. doi: 10.1038/nrgastro.2017.101

44. Yu S, Li B, Hao J, Shi X, Ge J, Xu H. Correlation of Hypoxia-Inducible Facto-1α and C-Reactive Protein With Disease Evaluation in Patients With Ulcerative Colitis. *Am J Transl Res* (2020) 12(12):7826–35.
45. Wang HQ, Zhang WH, Wang YQ, Geng XP, Wang MW, Fan YY, et al. Colonic Vitamin D Receptor Expression Is Inversely Associated With Disease Activity and Jumonji Domain-Containing 3 in Active Ulcerative Colitis. *World J Gastroenterol* (2020) 26(46):7352–66. doi: 10.3748/wjg.v26.i46.7352
46. Tan B, Li P, Lv H, Li Y, Wang O, Xing XP, et al. Vitamin D Levels and Bone Metabolism in Chinese Adult Patients With Inflammatory Bowel Disease. *J Dig Dis* (2014) 15(3):116–23. doi: 10.1111/1751-2980.12118
47. Shirwaikar Thomas A, Criss ZK, Shroyer NF, Abraham BP. Vitamin D Receptor Gene Single Nucleotide Polymorphisms and Association With Vitamin D Levels and Endoscopic Disease Activity in Inflammatory Bowel Disease Patients: A Pilot Study. *Inflamm Bowel Dis* (2021) 27(8):1263–9. doi: 10.1093/ibd/izaa292
48. Gubatan J, Rubin SJS, Bai L, Haileselassie Y, Levitte S, Balabanis T, et al. Vitamin D Is Associated With α4β7+ Immunophenotypes and Predicts Vedolizumab Therapy Failure in Patients With Inflammatory Bowel Disease. *J Crohns Colitis* (2021) 28:jjab114. doi: 10.1093/ecco-jcc/jjab114
49. Kobayashi T, Siegmund B, Le Berre C, Wei SC, Ferrante M, Shen B, et al. Ulcerative Colitis. *Nat Rev Dis Primers* (2020) 6(1):74. doi: 10.1038/s41572-020-0205-x
50. Bouma G, Strober W. The Immunological and Genetic Basis of Inflammatory Bowel Disease. *Nat Rev Immunol* (2003) 3(7):521–33. doi: 10.1038/nri1132
51. Stio M, Bonanomi AG, d'Albasio G, Treves C. Suppressive Effect of 1,25-Dihydroxyvitamin D3 and Its Analogues EB 1089 and KH 1060 on T Lymphocyte Proliferation in Active Ulcerative Colitis. *Biochem Pharmacol* (2001) 61(3):365–71. doi: 10.1016/s0006-2952(00)00564-5
52. Dulic S, Toldi G, Sava F, Kovács L, Molnár T, Milassin Á, et al. Specific T-Cell Subsets Can Predict the Efficacy of Anti-TNF Treatment in Inflammatory Bowel Diseases. *Arch Immunol Ther Exp (Warsz)* (2020) 68(2):12. doi: 10.1007/s00005-020-00575-5
53. Song J, Chen X, Cheng L, Rao M, Chen K, Zhang N, et al. D Receptor Restricts T Helper 2-Biased Inflammation in the Heart. *Cardiovasc Res* (2018) 114 (6):870–9. doi: 10.1093/cvr/cvy034
54. Dong B, Zhou Y, Wang W, Scott J, Kim K, Sun Z, et al. Vitamin D Receptor Activation in Liver Macrophages Ameliorates Hepatic Inflammation, Steatosis, and Insulin Resistance in Mice. *Hepatology* (2020) 71(5):1559–74. doi: 10.1002/hep.30937
55. Na YR, Stakenborg M, Seok SH, Matteoli G. Macrophages in Intestinal Inflammation and Resolution: A Potential Therapeutic Target in IBD. *Nat Rev Gastroenterol Hepatol* (2019) 16(9):531–43. doi: 10.1038/s41575-019-0172-4
56. Theill LE, Boyle WJ, Penninger JM. RANK-L and RANK: T Cells, Bone Loss, and Mammalian Evolution. *Annu Rev Immunol* (2002) 20:795–823. doi: 10.1146/annurev.immunol.20.100301.064753
57. Anderson DM, Maraskovsky E, Billingsley WL, Dougall WC, Tometsko ME, Roux ER, et al. A Homologue of the TNF Receptor and its Ligand Enhance T-Cell Growth and Dendritic-Cell Function. *Nature* (1997) 390(6656):175–9. doi: 10.1038/36593
58. Rabe H, Malmquist M, Barkman C, Östman S, Gjertsson I, Saalman R, et al. Distinct Patterns of Naive, Activated and Memory T and B Cells in Blood of Patients With Ulcerative Colitis or Crohn's Disease. *Clin Exp Immunol* (2019) 197(1):111–29. doi: 10.1111/cei.13294
59. Mitsialis V, Wall S, Liu P, Ordovas-Montanes J, Parmet T, Vukovic M, et al. Single-Cell Analyses of Colon and Blood Reveal Distinct Immune Cell Signatures of Ulcerative Colitis and Crohn's Disease. *Gastroenterology* (2020) 159(2):591–608.e10. doi: 10.1053/j.gastro.2020.04.074
60. de Souza HS, Fiocchi C. Immunopathogenesis of IBD: Current State of the Art. *Nat Rev Gastroenterol Hepatol* (2016) 13(1):13–27. doi: 10.1038/nrgastro.2015.186
61. Beswick L, Rosella O, Rosella G, Headon B, Sparrow MP, Gibson PR, et al. Exploration of Predictive Biomarkers of Early Infliximab Response in Acute Severe Colitis: A Prospective Pilot Study. *J Crohns Colitis* (2018) 12(3):289–97. doi: 10.1093/ecco-jcc/jjx146
62. Luo X, Davis JJ. Electrical Biosensors and the Label Free Detection of Protein Disease Biomarkers. *Chem Soc Rev* (2013) 42(13):5944–62. doi: 10.1039/c3cs60077g

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

ORIGINAL RESEARCH
published: 27 January 2022
doi: 10.3389/fimmu.2021.815121

Genome-Wide Association Study of Ustekinumab Response in Psoriasis

William T. Connell[1,2], *Julie Hong*[3] *and Wilson Liao*[3*]

[1] *Deparment of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA, United States,*
[2] *Insitute for Neurodegenerative Disease, University of California San Francisco, San Francisco, CA, United States,*
[3] *Department of Dermatology, University of California San Francisco, San Francisco, CA, United States*

Heterogeneous genetic and environmental factors contribute to the psoriasis phenotype, resulting in a wide range of patient response to targeted therapies. Here, we investigate genetic factors associated with response to the IL-12/23 inhibitor ustekinumab in psoriasis. To date, only HLA-C*06:02 has been consistently reported to associate with ustekinumab response in psoriasis. Genome-wide association testing was performed on the continuous outcome of percent change in Psoriasis Area Severity Index (PASI) at 12 weeks of ustekinumab therapy relative to baseline. A total of 439 European ancestry individuals with psoriasis were included [mean age, 46.6 years; 277 men (63.1%)]. 310 (70.6%) of the participants comprised the discovery cohort and the remaining 129 (29.4%) individuals comprised the validation cohort. Chromosome 4 variant rs35569429 was significantly associated with ustekinumab response at 12 weeks at a genome-wide significant level in the discovery cohort and replicated in the validation cohort. Of psoriasis subjects with at least one copy of the deletion allele of rs35569429, 44% achieved PASI75 (75% improvement in PASI from baseline) at week 12 of ustekinumab treatment, while for subjects without the deletion allele, 75% achieved PASI75 at week 12. We found that differences in treatment response increased when rs35569429 was considered alongside HLA-C*06:02. Psoriasis patients with the deletion allele of rs35569429 who were HLA-C*06:02 negative had a PASI75 response rate of 35% at week 12, while those without the deletion allele who were HLA-C*06:02 positive had a PASI75 response rate of 82% at week 12. Through GWAS, we identified a novel SNP that is potentially associated with response to ustekinumab in psoriasis.

Keywords: GWAS, psoriasis, ustekinumab, pharmacogenetics, precision medicine, pharmacogenomics

OPEN ACCESS

Edited by:
Xu-jie Zhou,
Peking University First Hospital, China

Reviewed by:
Then Fang Tsai,
National Taiwan University Hospital, Taiwan
Vinod Chandran,
University of Toronto, Canada

Correspondence:
Wilson Liao
wilson.liao@ucsf.edu

Specialty section:
This article was submitted to Autoimmune and Autoinflammatory Disorders, a section of the journal Frontiers in Immunology

Received: *15 November 2021*
Accepted: *27 December 2021*
Published: *27 January 2022*

Citation:
Connell WT, Hong J and Liao W (2022) Genome-Wide Association Study of Ustekinumab Response in Psoriasis. Front. Immunol. 12:815121. doi: 10.3389/fimmu.2021.815121

INTRODUCTION

Psoriasis is a common, chronic immune-mediated skin disease that affects at least 2% of the population worldwide (1). Psoriasis is associated with psoriatic arthritis, cardiovascular disease, metabolic syndrome, and other comorbidities, which makes effective management of psoriasis critical. Moderate-to-severe psoriasis is treated with phototherapy and systemic agents, including targeted biologic inhibitors of TNF-α, IL-12/23, IL-17, and IL-23. Patient responses to biologic therapy can vary widely, from poor overall response to gradual loss of therapeutic sensitivity (2). Response differences are largely influenced by patient weight and adherence, drug dose and bioavailability, and pharmacokinetic covariates, such as drug immunogenicity (3). The molecular heterogeneity of psoriasis may also contribute to differential therapeutic responses. However, there are no molecular biomarkers routinely used in clinical practice to facilitate selection of the therapies tailored to individual patients.

Ustekinumab is a fully humanized immunoglobulin monoclonal antibody targeting the p40 subunit shared by IL-12 and IL-23. Phase 3 clinical trials showed that treatment with ustekinumab results in 75% improvement in the Psoriasis Area and Severity Index (PASI75) in ~66% of patients after 12 weeks of therapy (4–6). Candidate gene studies have identified the HLA-C*06:02 allele as being associated with better ustekinumab responses in both European (7–9) and Chinese (10) patients with psoriasis. A meta-analysis of eight studies including 1048 psoriasis patients showed that HLA-C*06:02 positive patients had a median PASI75 response rate of 92% after 6 months of ustekinumab therapy compared to a median PASI75 response rate of 67% in the HLA-C*06:02 negative patients (11).

Here, we performed an unbiased genome-wide association study (GWAS) to evaluate if additional genetic factors were associated with ustekinumab response. We evaluated our findings across multiple response timepoints and in conjunction with HLA-C*06:02. Our findings highlight a potentially novel variant associated with ustekinumab response in psoriasis, which may further facilitate the development of precision medicine approaches.

MATERIALS AND METHODS

Study Population

This study involved analysis of individuals with moderate to severe psoriasis who participated in at least one of three placebo-controlled randomized clinical trials: PHOENIX I, PHOENIX II, and ACCEPT (4, 5, 12). Participants were originally approached for retrospective collection of DNA samples by investigators analyzing the association between the HLA-C*06:02 allele and response to IL-12/23 inhibition (7). In total, 439 patients of European descent were used to assess genetic associations between ustekinumab treatment and response.

The GWAS discovery cohort consisted of 310 individuals who were treated with 45mg (n=146) or 90mg (n=164) of ustekinumab for 40 weeks, with the lower or higher dose given according to body weight less than or greater than 100 kg, respectively. The validation cohort consisted of 129 trial participants who crossed-over from placebo to ustekinumab treatment at week 12 and continued ustekinumab for 16 weeks, again dose-stratified by body weight (45 mg: n=64; 90 mg: n=65). In both cohorts, ustekinumab was given with two loading doses 4 weeks apart and every 12 weeks thereafter (**Figure 1A**).

Response Variables

In the ustekinumab phase 3 clinical trials, the primary endpoint was achievement of PASI75 at week 12. PASI75 is a binary outcome converted from percent PASI improvement from

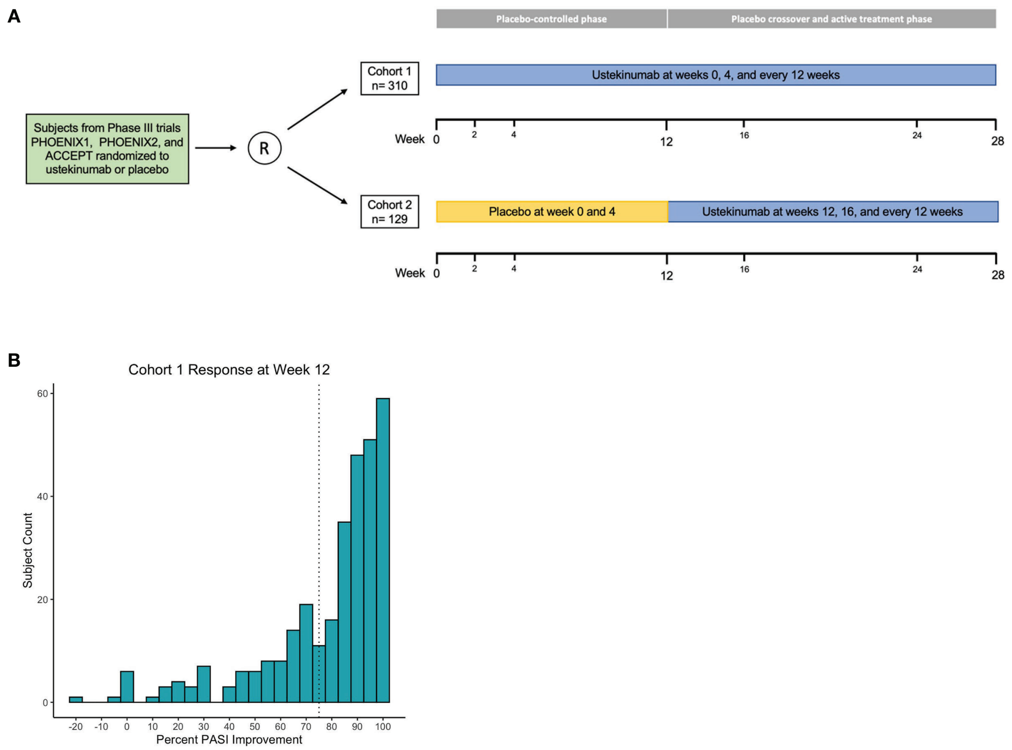


FIGURE 1 | Association analysis design and primary outcome. Phase 3 clinical trial comprise discovery and validation cohorts **(A)**. Histogram of cohort 1 percent PASI improvement at week 12; dashed line marks 75% improvement threshold **(B)**.

baseline. To maximize power for the GWAS, we focused on the continuous outcome measure of percent PASI improvement from baseline to 12 weeks after ustekinumab therapy. Phenotypic response to ustekinumab was recorded at weeks 2, 4, 12, 28, and 40 for the majority of patients in the discovery cohort (cohort 1). In order to validate findings, the placebo to ustekinumab cross-over patients acted as a validation cohort (cohort 2). PASI responses for cohort 2 were measured after 12 weeks of ustekinumab therapy compared to trial start.

Genome Wide Association Study

Genotyping was performed using Illumina HumanOmni2.5-8 v1.2 BeadChips. Imputation was performed using the Michigan Imputation Server (https://imputationserver.sph.umich.edu/index.html) (13). The 1000 Genomes Phase 3 data was used as a reference panel for imputation (14). Files were converted to PLINK (v1.9) format, which along with R (v3.5.1) and python (v3.7.4), was used for data manipulation, visualization, and association analysis. Quality control and population stratification was performed following methods outlined by Marees et al. (15). Single nucleotide polymorphisms (SNPs) and individuals with missingness greater than 2% were removed. Duplicate, non-biallelic, and poor imputation quality (R^2<0.7) SNPs were filtered. Non-autosomal SNPs with a low minor allele frequency (MAF<0.05) and significant deviation from Hardy-Weinberg equilibrium ($P<1\times10^{-6}$) were removed. In total 6,799,417 SNPs passed quality control, of which 1,696,820 were directly genotyped. Individuals with a heterozygosity rate +/-3 standard deviation from the mean were filtered, as well as the individual with the lowest call rate within a pair of cryptically related individuals ($\hat{\pi} > 0.2$). In total, 310 individuals (181 males, 129 females) passed quality control. The previously described quality control steps were applied to the 1000 Genomes Phase 3 data prior to merging with cohort data for population stratification. Multidimensional scaling (MDS) was applied to the merged genotype information. The presence of ethnic outliers was evaluated by qualitative alignment with the European superpopulation cluster along the top 2 MDS components. We included the top 10 MDS components as covariates in linear regression models for association testing.

Statistical Analysis

A threshold of $P<5\times10^{-8}$ was established in the discovery cohort to determine the associated markers for further replication. We took linkage-disequilibrium into account when interpreting multiple significant association results from the same region. Clumping was employed to greedily assign groups around index variants with $P<5\times10^{-6}$. Variants with an R^2>0.5 and less than 1MB away were assigned representation by the index variant. We modeled the additive effect of allele dosage with the quantitative phenotype of interest using linear regression. When considering cohort 1 index variants in replication analyses, a 2-sided *t*-test with $P<0.05$ was considered statistically significant. A two-sided normal test for proportions ($P<0.05$) was applied to assess PASI threshold achievement differences based on genotype. The combined cohort association study followed the same procedures outlined for analysis of discovery cohort results.

Power Analysis

We performed power calculations for the discovery and replication cohorts assuming an additive linear model for our quantitative trait of interest. Each power calculation was performed under consideration of the established type 1 error rates for the respective cohort (cohort 1 α, 5×10^{-8}; cohort 2 α, 5×10^{-2}). We examined power across a range of MAF (0.05-0.25) and effect sizes (ES) (1–9). The genpwr (v1.0.4) R package was used for all calculations.

RESULTS

In this study, we analyzed genetic data from two cohorts of psoriasis patients receiving ustekinumab. Following preprocessing and filtering for individuals of European genetic ancestry, the discovery cohort (cohort 1) totaled 310 individuals (181 males, 171 females) and the validation cohort (cohort 2) totaled 129 individuals (82 males, 47 females). The average PASI score at baseline was 18.6 for cohort 1 and 18.8 for cohort 2 (**Supplementary Table 1**). Power analysis revealed the discovery cohort had 1-β>0.75 for MAF>0.05 and ES>7. The replication cohort had 1-β>0.75 for MAF>0.05 and ES>5 (**Supplementary Figures 2A, B**). We used linear regression to perform genome-wide association testing on the percent improvement in PASI response at week 12 of ustekinumab therapy compared to baseline (**Figure 1B**). There was no correlation between age, BMI, and duration of the disease with the primary outcome of percent PASI improvement, and so these clinical variables were not included as covariates in the linear regression model (**Supplementary Figure 1**).

Genome-wide association testing of subjects in cohort 1 identified a single peak on chromosome 4 exceeding a genome-wide significance threshold of $P<5\times10^{-8}$ lead by rs35569429 (β, -19.84; 95% CI, -26.58 to -13.1; $P=1.98\times10^{-8}$) (**Figure 2A** and **Table 1**). Directly genotyped SNP rs11722643 was in high linkage disequilibrium with imputed SNP rs35569429 and achieved a similar level of significance (R^2, 0.9; β, -19.31; 95% CI, -26.33 to -12.29; $P=1.44\times10^{-7}$). To determine whether multiple SNPs contributed to the peak on chromosome 4, we performed conditional analysis on rs35569429. The conditional analysis completely attenuated the GWAS peak, indicating a single independent signal at this locus (**Figures 2B, C**). The major allele of rs35569429 is "G" while the minor allele is a single nucleotide deletion of G, denoted as "Del". Subjects with at least one minor allele were labeled as the deletion positive group (Del+, N=55), and subjects with zero minor alleles were labeled the deletion negative group (Del-, N=255). Only one subject was homozygous for the minor allele. To understand the impact of this SNP at various discrete levels of PASI response, we examined the proportions of Del- and Del+ individuals who achieved PASI50, PASI75, PASI90, and PASI100 at Week 12. We found that in the Del- group, 235/255 (92.2%) achieved PASI50, 191/255 (74.9%) achieved PASI75, 121/255 (47.5%) achieved PASI90, and 48/255 (18.8%) achieved PASI100 at Week 12. In the Del+ group, 39/55 (80.9%) achieved PASI50, 24/55 (43.6%) achieved

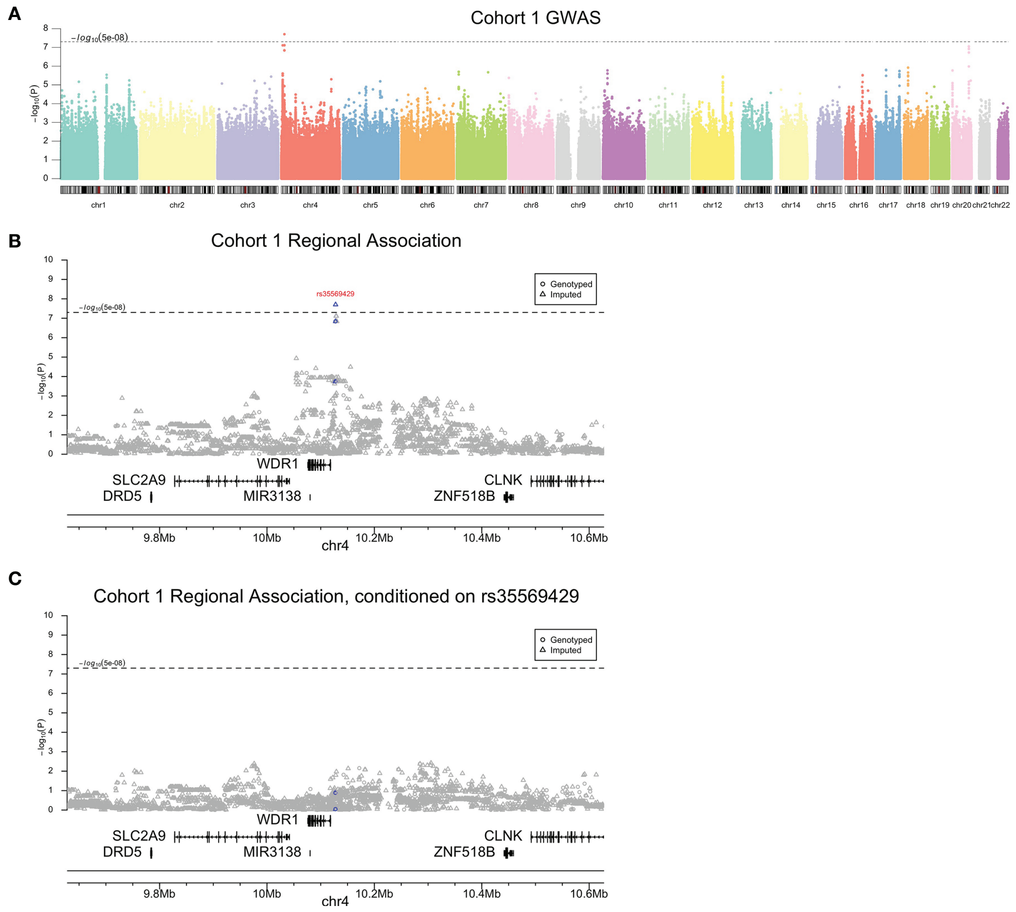


FIGURE 2 | Cohort 1 association analysis results. Genome-wide **(A)**, regional **(B)**, and conditional association **(C)** Manhattan plots. Blue indicates variants in high linkage disequilibrium (R^2>0.95) with rs35569429.

PASI75, 12/55 (21.8%) achieved PASI90, 5/55 (9.1%) achieved PASI100 at Week 12.

To further investigate the validity of rs35569429, we analyzed its association with PASI outcomes in cohort 1 at timepoints that were not part of the original GWAS analysis (i.e. timepoints other than week 12). We found that a greater proportion of individuals in the Del- group achieved PASI75 compared to the Del+ group at Week 2 (1.57% *vs* 0%), Week 4 (17.6% *vs* 10.9%), Week 24 (76.5% *vs* 61.8%), and Week 28 (73.3% *vs* 52.7%) (**Figure 3A**). Similarly, the Del- group had a higher proportion of individuals achieving PASI50, PASI90, and PASI100 than the Del+ group at weeks 2, 4, 24, and 28. The difference in PASI responses between Del- and Del+ groups were generally comparable if not greater than the difference in PASI responses between HLA-C*06:02 positive and HLA-C*06:02 negative individuals (**Figure 3B**), where HLA-C*06:02 represents a previously well-validated locus associated with ustekinumab response (11). For comparison, in cohort 1, a linear regression of PASI response at week 12 for HLA-C*06:02, using 10 MDS components as co-variates, yielded β=0.7418 and *P*=0.0093.

We next investigated the association of rs35569429 with response to ustekinumab in an independent cohort 2. We found the same direction of effect at week 12 for rs35569429 (*β*, -6.71; 95% CI, -13.13 to -0.30; *P*=0.042) (**Table 1**). In the Del- group, 102/106 (94.5%) subjects achieved PASI50, 81/106 (76.4%) subjects achieved PASI75, 45/106 (42.5%) achieved

TABLE 1 | Cohort 1, 2 and combined association analysis results.

	SNP	MAF	β	*P* value
Cohort 1	rs35569429	0.090	-19.84	1.98E-08
Cohort 2	rs35569429	0.097	-6.71	0.042
Cohort 1 + 2 Combined Analysis	rs35569429	0.092	-15.83	2.42E-09

MAF, mean allele frequency.

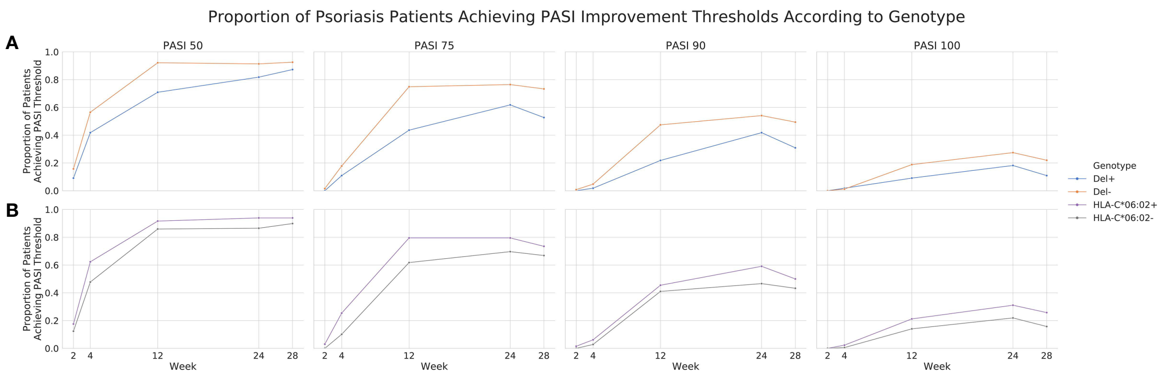


FIGURE 3 | Proportion of psoriasis patients achieving PASI thresholds according to genotype in cohort 1. PASI 50, 75, 90 and 100 achievement across weeks 2, 4, 12, 24 and 28 for rs35569429 **(A)** and HLA-C*06:02 **(B)** genotypes.

PASI90, and 26/106 (24.5%) achieved PASI100 at Week 12. In the Del+ group, 20/23 (87.0%) subjects achieved PASI50, 13/23 (56.5%) achieved PASI75, 9/23 (39.1%) achieved PASI90, and 2/23 (8.7%) achieved PASI100 at Week 12. Association testing for rs35569429 in cohort 1 and cohort 2 combined at week 12 yielded a genome-wide significant result (β, -15.83; 95% CI, -20.72 to -10.74; P=2.42×10^{-9}). We ran a sensitivity analysis on the full sample of cohorts 1 and 2 combined at week 12. We observed the expected genome-wide significant peak at rs35569429, with the most significant SNP being rs11722643, which is in high linkage disequilibrium with rs35569429 (R^2, 0.88; β, -16.64; 95% CI, -22.04 to -11.25; P=3.25×10^{-9}). We also observed a single additional genome-wide significant loci on chromosome 14, which could not be further confirmed (rs994384156; β, -14.94; 95% CI, -20.02 to -9.86; P=1.58×10^{-8}). We also conducted a separate GWAS on ustekinumab response at week 24 and did not identify any genome-wide significant SNPs.

Finally, we explored how the combination of rs35569429 and HLA-C*06:02 affects PASI75 response in cohort 1 and 2 at week 12, since HLA-C*06:02 is an allele previously established to be associated with a more favorable responses to ustekinumab in psoriasis (11). In cohort 1 at week 12, 82.4% Del-/HLA-C*06:02+ individuals achieved PASI75 compared to 68.8% in Del-/HLA-C*06:02-, 61.1% in Del+/HLA-C*06:02+, and 35.1% in Del+/HLA-C*06:02- (**Figure 4**). In cohort 2 at week 12, 88.6% Del-/HLA-C*06:02+ individuals achieved PASI75 compared to 79.2% in Del-/HLA-C*06:02-, 72.7% in Del+/HLA-C*06:02+, and 50.0% in Del+/HLA-C*06:02-. In cohort 1 and cohort 2 combined at week 12, 84.4% Del-/HLA-C*06:02+ individuals achieved PASI75 compared to 71.6% in Del-/HLA-C*06:02-, 65.5% in Del+/HLA-C*06:02+, and 38.8% in Del+/HLA-C*06:02. The effects of rs35569429 and HLA-C*06:02 were independent from each other, as an interaction analysis that included an interaction term between rs35569429 and HLA-C*06:02 was not significant (P=0.729).

DISCUSSION

This genetic association study found a genome-wide significant association between intergenic variant rs35569429 and response to ustekinumab for the treatment of moderate to severe psoriasis. In our primary association analysis, absence of the minor allele (Del-) was significantly associated with a larger PASI improvement at 12 weeks from baseline. More favorable PASI

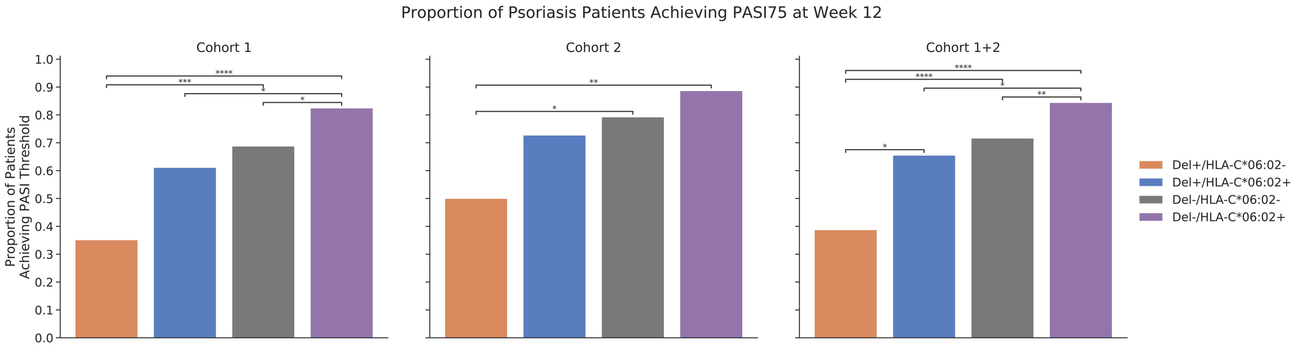


FIGURE 4 | Proportion of psoriasis patients achieving PASI75 at week 12. *P<=5×10^{-2}; **P<=1×10^{-2}; ***P<=1×10^{-3}; ****P<=1×10^{-4}.

responses in Del- individuals compared to Del+ individuals were also observed at weeks 2, 4, 24, and 28. The association of rs35569429 with ustekinumab response was validated in an independent cohort of psoriasis patients. Conditional analysis revealed a single independent signal at the locus of interest.

rs35569429 is characterized by a G deletion minor allele. This variant is located in an intergenic region 9 kB upstream of *WDR1*. Functional analysis by GeneHancer Regulatory Elements strongly associates a 10.6 kB region (GH04J010114) 1.2 kB downstream of this variant with promoter/enhancer activity influencing proximal protein coding genes *WDR1* and *SLC2A9* (16). The WDR1 protein is involved in actin filament disassembly, a critical process of cytoskeleton dynamics, especially in highly motile and interacting immune cells (17). Impaired actin dynamics as a result of WDR1 deficiency have been causally linked to primary immunodeficiencies and autoinflammatory phenotypes (18, 19). SLC2A9 is a transporter mainly expressed in the kidneys and primarily involved in urate reabsorption. Mutations of *SLC2A9* lead to poor reabsorption and Renal Hypouricemia type-2, as caused by increased urate excretion (20). Future studies are needed to fine-map the causal and functional SNPs in linkage disequilibrium with rs35569429.

Stratification of ustekinumab responses was greatest when rs35569429 was considered in combination with HLA-C*06:02. Individuals who were Del-/HLA-C*06:02+ achieved PASI75 84.4% of the time, while those were Del+/HLA-C*06:02- achieved PASI75 38.8% of the time, a more than two-fold difference.

Pharmacogenomics continues to play an increasingly important role in precision medicine for dermatology. In 2018, five dermatologic drugs had clinically actionable pharmacogenomic tags that either require or advise testing of genomic biomarkers before treatment (21). Single FDA-approved biomarkers currently dominate this list; however, multi-gene marker panels will continue to gain importance for informing clinical decisions. Understanding the role of multiple SNPs in disease pathogenesis is important in advancing precision medicine.

Conclusions from this study are limited due to the moderate sample size of the discovery and replication cohorts; our study was not powered for detection of small to moderate effects. Given the polygenicity of complex autoimmune diseases such as psoriasis, in the future, prospective design of large study cohorts is essential for thorough investigation of the biology contributing to therapeutic response. In general, validation in additional, independent cohorts will provide evidence with respect to the genomic signals discovered herein. Furthermore, the index SNP rs35569429 requires further investigation to identify the causal variant(s) associated with this locus and further characterization of functional effects on psoriatic response to ustekinumab.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: (https://doi.org/10.6084/m9.figshare.17009930).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Boards of each clinical trial site participating in ustekinumab phase 3 studies. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

WL conceived and supervised the project. WC performed GWAS. WC and JH performed data analysis, prepared, and wrote the manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We would like to acknowledge Anne Fourie, Xuejun Liu, Amy Hart, and Dawn Waterworth at Janssen Research & Development, LLC for providing access to the data and for providing analysis feedback. We would also like to acknowledge Joanne Nititham for assistance with data handling. Janssen provided the genotyping and clinical response data for the study from ustekinumab pivotal trials, but was not involved in the funding, study design, data analysis, data interpretation, manuscript writing, or decision to submit the manuscript for publication. WL had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.815121/full#supplementary-material

REFERENCES

1. Parisi R, Symmons DPM, Griffiths CEM, Ashcroft DM Identification and Management of Psoriasis and Associated ComorbidiTy (IMPACT) Project Team. Global Epidemiology of Psoriasis: A Systematic Review of Incidence and Prevalence. *J Invest Dermatol* (2013) 133(2):377–85. doi: 10.1038/jid.2012.339
2. Warren RB, Smith CH, Yiu ZZN, Ashcroft DM, Barker JNWN, Burden AD, et al. Differential Drug Survival of Biologic Therapies for the Treatment of Psoriasis: A Prospective Observational Cohort Study From the British Association of Dermatologists Biologic Interventions Register (BADBIR). *J Invest Dermatol* (2015) 135(11):2632–40. doi: 10.1038/jid.2015.208
3. Mahil SK, Capon F, Barker JN. Update on Psoriasis Immunopathogenesis and Targeted Immunotherapy. *Semin Immunopathol* (2016) 38(1):11–27. doi: 10.1007/s00281-015-0539-8
4. Leonardi CL, Kimball AB, Papp KA, Yeilding N, Guzzo C, Wang Y, et al. Efficacy and Safety of Ustekinumab, a Human Interleukin-12/23 Monoclonal

Antibody, in Patients With Psoriasis: 76-Week Results From a Randomised, Double-Blind, Placebo-Controlled Trial (PHOENIX 1). *Lancet Lond Engl* (2008) 371(9625):1665–74. doi: 10.1016/S0140-6736(08)60725-4

5. Papp KA, Langley RG, Lebwohl M, Krueger GG, Szapary P, Yeilding N, et al. Efficacy and Safety of Ustekinumab, a Human Interleukin-12/23 Monoclonal Antibody, in Patients With Psoriasis: 52-Week Results From a Randomised, Double-Blind, Placebo-Controlled Trial (PHOENIX 2). *Lancet Lond Engl* (2008) 371(9625):1675–84. doi: 10.1016/S0140-6736(08)60726-6
6. Young MS, Horn EJ, Cather JC. The ACCEPT Study: Ustekinumab Versus Etanercept in Moderate-to-Severe Psoriasis Patients. *Expert Rev Clin Immunol* (2011) 7(1):9–13. doi: 10.1586/eci.10.92
7. Li K, Huang CC, Randazzo B, Li S, Szapary P, Curran M, et al. HLA-C*06:02 Allele and Response to IL-12/23 Inhibition: Results From the Ustekinumab Phase 3 Psoriasis Program. *J Invest Dermatol* (2016) 136(12):2364–71. doi: 10.1016/j.jid.2016.06.631
8. Talamonti M, Galluzzo M, van den Reek JM, de Jong EM, Lambert JLW, Malagoli P, et al. Role of the HLA-C*06 Allele in Clinical Response to Ustekinumab: Evidence From Real Life in a Large Cohort of European Patients. *Br J Dermatol* (2017) 177(2):489–96. doi: 10.1111/bjd.15387
9. Talamonti M, Galluzzo M, Chimenti S, Costanzo A. HLA-C*06 and Response to Ustekinumab in Caucasian Patients With Psoriasis: Outcome and Long-Term Follow-Up. *J Am Acad Dermatol* (2016) 74(2):374–5. doi: 10.1016/j.jaad.2015.08.055
10. Chiu H-Y, Wang T-S, Chan C-C, Cheng Y-P, Lin S-J, Tsai T-F. Human Leucocyte Antigen-Cw6 as a Predictor for Clinical Response to Ustekinumab, an Interleukin-12/23 Blocker, in Chinese Patients With Psoriasis: A Retrospective Analysis. *Br J Dermatol* (2014) 171(5):1181–8. doi: 10.1111/bjd.13056
11. van Vugt LJ, van den Reek JMPA, Hannink G, Coenen MJH, de Jong EMGJ. Association of HLA-C*06:02 Status With Differential Response to Ustekinumab in Patients With Psoriasis: A Systematic Review and Meta-Analysis. *JAMA Dermatol* (2019) 01155(6):708–15. doi: 10.1001/jamadermatol.2019.0098
12. Griffiths CEM, Strober BE, van de Kerkhof P, Ho V, Fidelus-Gort R, Yeilding N, et al. Comparison of Ustekinumab and Etanercept for Moderate-to-Severe Psoriasis. *N Engl J Med* (2010) 362(2):118–28. doi: 10.1056/NEJMoa0810652
13. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-Generation Genotype Imputation Service and Methods. *Nat Genet* (2016) 48 (10):1284–7. doi: 10.1038/ng.3656
14. Siva N. 1000 Genomes Project. *Nat Biotechnol* (2008) 26(3):256–6. doi: 10.1038/nbt0308-256b
15. Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, et al. A Tutorial on Conducting Genome-Wide Association Studies: Quality Control and Statistical Analysis. *Int J Methods Psychiatr Res* (2018) 27(2): e1608. doi: 10.1002/mpr.1608
16. Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, et al. GeneHancer: Genome-Wide Integration of Enhancers and Target Genes in GeneCards. *Database J Biol Database Curation* (2017) 2017. doi: 10.1093/database/bax028
17. Pfajfer L, Mair NK, Jiménez-Heredia R, Genel F, Gulez N, Ardeniz Ö, et al. Mutations Affecting the Actin Regulator WD Repeat-Containing Protein 1 Lead to Aberrant Lymphoid Immunity. *J Allergy Clin Immunol* (2018) 142 (5):1589–1604.e11. doi: 10.1016/j.jaci.2018.04.023
18. Standing ASI, Malinova D, Hong Y, Record J, Moulding D, Blundell MP, et al. Autoinflammatory Periodic Fever, Immunodeficiency, and Thrombocytopenia (PFIT) Caused by Mutation in Actin-Regulatory Gene WDR1. *J Exp Med* (2017) 214(1):59–71. doi: 10.1084/jem.20161228
19. Kuhns DB, Fink DL, Choi U, Sweeney C, Lau K, Priel DL, et al. Cytoskeletal Abnormalities and Neutrophil Dysfunction in WDR1 Deficiency. *Blood* (2016) 128(17):2135–43. doi: 10.1182/blood-2016-03-706028
20. Ruiz A, Gautschi I, Schild L, Bonny O. Human Mutations in SLC2A9 (Glut9) Affect Transport Capacity for Urate. *Front Physiol* (2018) 9:476. doi: 10.3389/fphys.2018.00476
21. Do LHD, Maibach H. Pharmacogenomics/updated for Precision Medicine in Dermatology. *J Dermatol Treat* (2019) 30(4):410–3. doi: 10.1080/09546634.2018.1527434

Conflict of Interest: WL has received research grant funding from Abbvie, Amgen, Janssen, Leo, Novartis, Pfizer, Regeneron, and TRex Bio.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers
in Immunology

ORIGINAL RESEARCH
published: 02 March 2022
doi: 10.3389/fimmu.2022.835760

Combined Single Cell Transcriptome and Surface Epitope Profiling Identifies Potential Biomarkers of Psoriatic Arthritis and Facilitates Diagnosis *via* Machine Learning

Jared Liu[1], Sugandh Kumar[1], Julie Hong[1], Zhi-Ming Huang[1], Diana Paez[2], Maria Castillo[2], Maria Calvo[2], Hsin-Wen Chang[1], Daniel D. Cummins[1], Mimi Chung[1], Samuel Yeroushalmi[1], Erin Bartholomew[1], Marwa Hakimi[1], Chun Jimmie Ye[2,3,4,5,6,7], Tina Bhutani[1], Mehrdad Matloubian[2,8], Lianne S. Gensler[2] and Wilson Liao[1,3]*

[1] Department of Dermatology, University of California at San Francisco, San Francisco, CA, United States, [2] Division of Rheumatology, Department of Medicine, University of California at San Francisco, San Francisco, CA, United States, [3] Institute for Human Genetics, University of California at San Francisco, San Francisco, CA, United States, [4] Department of Epidemiology and Biostatistics, University of California at San Francisco, San Francisco, CA, United States, [5] Institute of Computational Health Sciences, University of California at San Francisco, San Francisco, CA, United States, [6] Parker Institute for Cancer Immunotherapy, San Francisco, CA, United States, [7] Chan Zuckerberg Biohub, San Francisco, CA, United States, [8] Rosalind Russell/Ephraim P Engleman Rheumatology Research Center, University of California at San Francisco, San Francisco, CA, United States

Early diagnosis of psoriatic arthritis (PSA) is important for successful therapeutic intervention but currently remains challenging due, in part, to the scarcity of non-invasive biomarkers. In this study, we performed single cell profiling of transcriptome and cell surface protein expression to compare the peripheral blood immunocyte populations of individuals with PSA, individuals with cutaneous psoriasis (PSO) alone, and healthy individuals. We identified genes and proteins differentially expressed between PSA, PSO, and healthy subjects across 30 immune cell types and observed that some cell types, as well as specific phenotypic subsets of cells, differed in abundance between these cohorts. Cell type-specific gene and protein expression differences between PSA, PSO, and healthy groups, along with 200 previously published genetic risk factors for PSA, were further used to perform machine learning classification, with the best models achieving AUROC ≥ 0.87 when either classifying subjects among the three groups or specifically distinguishing PSA from PSO. Our findings thus expand the repertoire of gene, protein, and cellular biomarkers relevant to PSA and demonstrate the utility of machine learning-based diagnostics for this disease.

Keywords: psoriatic arthritis, psoriasis, CITE-seq, machine learning, diagnostic test, single cell

OPEN ACCESS

Edited by:
Xu-jie Zhou,
Peking University First Hospital, China

Reviewed by:
Eva Reali,
University of Milano-Bicocca, Italy
Iannis E Adamopoulos,
Beth Israel Deaconess Medical Center and Harvard Medical School, United States

****Correspondence:***
Wilson Liao
wilson.liao@ucsf.edu

Specialty section:
This article was submitted to Autoimmune and Autoinflammatory Disorders, a section of the journal Frontiers in Immunology

Received: *14 December 2021*
Accepted: *03 February 2022*
Published: *02 March 2022*

Citation:
Liu J, Kumar S, Hong J, Huang Z-M, Paez D, Castillo M, Calvo M, Chang H-W, Cummins DD, Chung M, Yeroushalmi S, Bartholomew E, Hakimi M, Ye CJ, Bhutani T, Matloubian M, Gensler LS and Liao W (2022) Combined Single Cell Transcriptome and Surface Epitope Profiling Identifies Potential Biomarkers of Psoriatic Arthritis and Facilitates Diagnosis via Machine Learning. Front. Immunol. 13:835760. doi: 10.3389/fimmu.2022.835760

INTRODUCTION

Psoriatic arthritis (PSA) is an inflammatory rheumatic disease that can affect the peripheral joints, axial joints, and entheses. PSA largely occurs in association with the skin disease psoriasis (PSO), with roughly a third of individuals with PSO developing PSA (1). Early detection of PSA in PSO patients is an important determinant of clinical outcome and patient long-term quality of life (2) but can be challenging due to the heterogeneous presentation of PSA, with only subclinical manifestations at early stages of disease (3).

The ongoing effort to develop better molecular diagnostics for PSA has identified genetic polymorphisms, primarily in major histocompatibility complex and IL-17/IL-23 signaling loci that contribute to PSA risk in PSO patients (4, 5), as well as disease-relevant immune cells within the inflamed synovium of affected joints. These include both adaptive and innate cell types that have a common inflammatory and IL-17-secreting role in pathogenesis and are significantly expanded in the synovium (6). Within peripheral blood, some cell types have also been reported to be perturbed in PSA patients, and while some studies have reported serum biomarkers for distinguishing PSA from PSO (7, 8), a more recent study found similar serum proteomes among PSO patients with and without PSA (9).

In this study, we searched for biomarkers of PSA within the circulating immune cell population by jointly measuring transcriptomic and cell surface protein expression of peripheral blood immune cells at the single cell level. Our data reveal PSA-associated differences in the abundance of phenotypic cell clusters within specific adaptive and innate immune subsets. We further examine disease-associated RNA and protein markers found in this analysis, along with genotype data from PSA-associated polymorphisms, developing a machine-learning-based diagnostic for distinguishing between PSA and PSO.

MATERIALS AND METHODS

Patient Recruitment and Sampling

Patients with PSO were enrolled from the dermatology clinics at the University of California San Francisco (UCSF), with a board-certified dermatologist confirming the clinical diagnosis of plaque psoriasis. Patients with PSA were assessed by a board-certified rheumatologist and diagnosed with PSA according to CASPAR criteria. Patients with psoriasis who reported symptoms of joint pain, but who did not meet CASPAR criteria, were assigned the label of PSX. Healthy controls, who did not have any inflammatory skin disease or autoimmune disease, were enrolled from the San Francisco Bay Area. All subjects gave written, informed consent under IRB approval 10-02830 from the University of California San Francisco. Detailed patient information is provided in **Supplementary Table 1**. Peripheral blood was collected from each subject in Vacutainer ACD tubes. PBMCs were isolated using a standard Ficoll method and stored in liquid nitrogen.

Sample and Library Preparation

Single Cell Libraries

500 µL thawed PBMCs from each subject were added to 10 mL EasySep (StemCell Technologies, Cat. 20144) and centrifuged (300G, 5 min, room temperature). Extracellular nucleic acids were digested by resuspending cell pellets in 1 mL of buffer made from 18 mL EasySep and 21 µL Benzonase Nuclease (MilliporeSigma, Cat. 70664) and incubating (15 min, room temperature). Nuclease-treated cell-suspensions were then filtered through a 40 µm Flowmi Cell Strainer (Bel-Art, Cat. H13680-0040), centrifuged (300G, 5 min, room temperature), and finally resuspended in 100 µL EasySep buffer. Cell counting was performed using a Countess I FL Automated Cell Counter (Thermo Fisher Scientific) on 1:100 dilutions of final cell suspensions stained with 0.4% trypan blue.

Cell Surface Staining

Antibody staining of cell surface proteins was performed according to the Totalseq-A protocol (https://www.biolegend.com/en-us/protocols/totalseq-a-antibodies-and-cell-hashing-with-10x-single-cell-3-reagent-kit-v3-3-1-protocol) with modifications as follows.

A pooled suspension containing 2×10^6 cells from 20 subjects at a time (~100,000 per subject) was centrifuged (300G, 5 min, 4°C) and resuspended in 100 µL Cell Staining Buffer (BioLegend, Cat. 420201) and incubated (10 min, 4°C) with 10 µL Human TruStain FcX^{TM} Fc Blocking Solution (BioLegend, Cat. 422301). Cells suspensions were then stained (30 min, 4°C) with 100 µL TotalSeq antibody cocktail for feature barcoding of cell surface proteins (**Supplementary Table 2**) and divided into two 105 µL aliquots. Each aliquot was washed 3 times by resuspending in 15 mL Cell Staining Buffer and centrifuging (300G, 5 min, 4°C). Aliquots of washed cells were then resuspended in 150 µL 10% FBS in PBS to obtain a concentration of 1×10^6 cells/mL, recombined, and filtered again with a 40 µm Flowmi Cell Strainer. Cell viability was measured with 10 µL of filtered cells by adding 10 µL 0.4% Trypan Blue and manually counting with a hemocytometer.

Cell density was adjusted to 2,500 cells/µL and run on the Chromium Controller (10X Genomics) using the Single Cell 3' v3.1 Assay (10X Genomics) with a target of 50,000 cells per reaction.

Library Preparation

Gene expression cDNA libraries were prepared according to the manufacturer's instructions (https://assets.ctfassets.net/an68im79xiti/1eX2FPdpeCgnCJtw4fj9Hx/7cb84edaa9eca04b607f9193162994de/CG000204_ChromiumNextGEMSingleCell3_v3.1_Rev_D.pdf), with 12 cycles of PCR amplification.

Libraries for antibody-derived tags (ADT) from feature barcoding antibodies were prepared by repeating size purification on the supernatant obtained from the prior size purification of gene expression cDNA libraries (Step 2.3.d in the manufacturer's instructions above), using a 7:8 volumetric ratio of 2.0X SPRIselect reagent (Beckman Coulter, Cat# B23317) to sample. Indexing amplification was performed using Kapa Hifi

HotStart ReadyMix (Kapa Biosystems, Cat# KK2601) and TruSeq Small RNA RPI primers (Illumina) with the following thermocycling conditions (1): 98°C, 2 min (2); 15 × (98°C, 20 sec; 60°C, 30 sec; 72°C, 20 sec) (3); 72°C, 5 min. Size purification was then repeated on amplified libraries using a 5:6 volumetric ratio of 1.2X SPRIselect reagent to sample.

Libraries were quantified using a Bioanalyzer 2100 (Agilent) and sequenced on a Novaseq 6000 (Illumina).

Genotyping

DNA for genotyping was extracted from whole blood using the DNeasy blood and tissue kit (Qiagen, Cat. 69504). Extracted DNA was genotyped on the Affymetrix UK Biobank Axiom Array (ThermoFisher) using a GeneTitan Multi-Channel Instrument (Applied Biosystems).

Genotype Data Processing

SNPs were called using Analysis Power Tools 2.10.2.2 (Affymetrix, https://www.affymetrix.com/support/developer/powertools/changelog/index.html). The resulting genotype.vcfs were scanned with 'snpflip' (https://github.com/biocore-ntnu/snpflip) using the GRCh37 build of the human genome reference sequence maintained by the University of California, Santa Cruz (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/bigZips/hg19.fa.gz) to identify reversed and ambiguous-stranded SNPs, which were flipped and removed (respectively) using Plink 1.90 (http://pngu.mgh.harvard.edu/purcell/plink/) (10), and the remaining sites were sorted using Plink 2.00a3LM (www.cog-genomics.org/plink/2.0/) (11). This SNP data was then augmented with additional sites imputed by the Michigan Imputation Server (https://imputationserver.sph.umich.edu) (1000G Phase 3 v5 GRCh37 reference panel, rsqFilter off, Eagle v2.4 phasing, EUR population). SNP positions were translated to GRCh38 coordinates using the 'LiftoverVcf' command of Picard 2.23.3 (http://broadinstitute.github.io/picard/). Finally, Vcftools 0.1.13 (12) was used to exclude non-exonic SNPs and SNPs with minor allele frequency < 0.05.

Single Cell Data Processing

Raw RNA and ADT fastqs for each Chromium library were respectively aligned to the GRCh38 human genome reference and the antibody-tag reference (**Supplementary Table 2**) using Cell Ranger 3.1.0 (10X Genomics) with default settings to obtain RNA and matched ADT (if available) count matrices for all barcodes representing non-empty droplets.

Cell Demultiplexing, Doublet Removal, and Annotation

Within each RNA count matrix, the subject of origin for all droplet barcodes was determined by using 'demuxlet' (13), as implemented in the 'popscle' suite (https://github.com/statgen/popscle), with imputation-augmented exonic SNP genotypes described above, and doublets detected between different individuals were excluded. The count matrices for each Chromium library were then loaded into R for analysis using the 'Seurat' 4.0.3 (14) R package, and the 'DoubletDecon' 1.1.6 R package (15) was used to further remove doublets formed by different cells within the same individual.

QC and Cell Annotation

Cell type annotation was performed by integrative mapping of annotations from a previously published dataset of 161,764 healthy PBMCs (14) onto our dataset. Specifically, we used the 'TransferData' Seurat function according to the Seurat protocol (https://satijalab.org/seurat/reference/transferdata) to transfer annotations for 30 distinct cell types from the 'predicted.celltype.l2' metadata variable.

We performed filtering of cells based on both RNA and ADT data by retaining cells with total RNA unique molecular identifiers (UMIs) between 500 and 10,000, total RNA features ≥ 200, percent mitochondrial and ribosomal protein reads in RNA ≤ 15% and 60% (respectively), total ADT features ≤ 260, and percent ADT reads mapping to 9 isotype control antibodies < 2%. In the RNA matrices of the resulting data, we further removed features (genes) with no detectable UMIs across the cells of all matrices. These matrices were finally merged into a combined matrix of RNA data for all cells. In the ADT matrices, we further removed features corresponding to the 9 isotype controls and 15 features observed to have expression inconsistent with annotated cell types (**Supplementary Table 2**). Lastly, we observed that a single healthy subject was represented by only 4 cells after filtering. These cells were excluded from later analysis.

ADT Imputation and UMAP Generation

ADT expression was estimated for cells with measured RNA but not ADT according to the Seurat reference mapping protocol (https://satijalab.org/seurat/articles/multimodal_reference_mapping.html), and unless otherwise noted, all function names described here belong to the Seurat package. Briefly, the integrated dataset above was split into the subset of cells with ADT measurements (reference subset) and the subset of cells without ADT measurements (query subset). RNA expression normalization and scaling were performed using 'SCTransform' on both subsets, adjusting for the number of features and total counts in each cell *via* the 'vars.to.regress' parameter. ADT expression normalization for the reference subset was performed using the centered log ratio (CLR), followed by mean centering and scaling. For the reference subset, PCA was then run for both the SCTransformed RNA (SCT) expression and the ADT expression, and a weighted nearest-neighbor network for the reference subset was calculated from the first 30 and 18 PCs for SCT and ADT, respectively, using the 'FindMultiModalNeighbors' function. Next, SCT from the reference subset was transformed again using supervised PCA (via the 'RunSPCA' function) to identify the principal components that best capture the combined RNA and ADT expression variation represented by the weighted nearest-neighbor network.

The first 50 components of this transformation were then used to identify anchors between the reference subset and the SCT of the query subset using the 'FindTransferAnchors' function. Finally, imputed ADT (ADTimp) data for the query subset was calculated using the 'TransferData' function. A weighted nearest-neighbor network was calculated using both SCT and ADTimp according to the Seurat protocol (https://satijalab.org/seurat/articles/weighted_nearest_neighbor_analysis.html).

Intra-Cell Type Differential Feature Analysis

To identify differentially expressed genes (DEGs) and proteins (DEPs), the Seurat object containing ADT and RNA expression from the QC'd dataset (see section QC and Annotation above) was subsetted by annotated cell type using 'SplitObject'. For each resulting Seurat object containing cells of a particular type, we performed normalization on RNA expression using SCTransform, again adjusting for processing batch ('Run' metadata variable) within each cell type (using the 'vars.to.regress' parameter of SCTransform). Differential gene expression between disease statuses as well as between clusters (see section '*Intra-cell type clustering*') was then calculated on SCTransform-normalized counts using the negative binomial test (test.use = "negbinom" in Seurat). Genes with both Bonferroni-corrected p-value < 0.05 and absolute log fold change > 0.25 were considered significant. Differential protein analysis was performed similarly, except with the Wilcoxon test (test.use = "wilcox" in Seurat) on CLR-normalized, mean-centered and scaled ADT data (within the 'scale.data' slot of the Seurat object) only for cells with measured (i.e. non-imputed) ADT data.

Cell Type Proportion Comparison

To detect statistical differences in the frequencies of each annotated cell type between cohorts, we calculated, for each cell type, the proportion of cells of that cell type in each subject out of the total number of cells in the subject, and the Kruskall-Wallis test ('kruskal.test' in R) was used to determine whether significant cell proportion differences existed between any cohort of subjects. For cell types with FDR-adjusted Kruskall-Wallis p-values < 0.05, we then performed Wilcoxon tests ('wilcox.test' in R) to identify significant (unadjusted p-value < 0.05) differences in cell proportions between cohorts. The same method was used to test for differences in the proportions of subclusters within cell types.

Intra-Cell Type Clustering

To identify phenotypic clusters within cell types, the RNA expression data for a cell type was first corrected for batch effects by first subsetting the raw count matrix by the cells within each sequencing batch. SCTransform was run individually for each count matrix, and the resulting SCT expression matrices were reintegrated into a single matrix (see section '*Data integration*'). PCA was performed on the integrated SCT matrix, and the first 30 PCs were used to construct a shared nearest-neighbor network using the 'FindNeighbors' function. The network was then used to identify clusters with the 'FindClusters' function, using a resolution of 0.6. UMAPs were also generated from the first 30 PCs using the 'RunUMAP' function.

Data Integration

Integration of SCT expression data from two or more single-cell datasets was performed according to the Seurat data integration protocol (https://satijalab.org/seurat/articles/integration_introduction.html#performing-integration-on-datasets-normalized-with-sctransform-1). Briefly, 'SelectIntegration Features' was used to select a common set of 3,000 genes most consistently variable among the individual SCT matrices, and 'Prep SCTIntegration' was then used to prepare reduced SCT expression matrices for just these genes. PCA was calculated for each reduced SCT matrix using 'RunPCA', and the first 50 principal components of this transformation were used to identify transcriptionally similar cells between each pair of reduced SCT matrices using 'FindIntegrationAnchors', with 'reduction' set to 'rpca'. Finally, an integrated SCT matrix was calculated using 'IntegrateData'.

Machine Learning Model Development

Input data for classifying each subject in PSA, PSO, healthy, and PSX cohorts was prepared by calculating the mean of the normalized, centered, and scaled expression of each feature in the set of cell type-specific differentially-expressed genes and proteins (found between PSA and healthy, PSO and healthy, and PSA and PSO groups; see section '*Intra-cell type differential feature analysis*') for all cells of the corresponding cell type in a given subject. The feature expression data for healthy, PSA, and PSO subjects (N=81) were then divided into a training set, n=58 (healthy=21, PSA=20, PSO=17) and a test set, n=23 (healthy=8, PSA=8, and PSO=7) to achieve a training:test ratio of 70:30.

We first performed ensemble-based feature selection using the EFS-MI method (16) where subsets of the starting feature set predicted to be informative by four different ML algorithms (Feed Forward and Backward selection, Recursive RF, SVMRadial, and NNET) were combined and sorted by prediction potential classification rank. We selected the top twenty features to train eleven ML algorithms such as linear, non-linear, and ensemble provided by the 'caret' R package, assessing classification performance using accuracy and kappa. To avoid overfitting and reduce the noise of random fitting models, we employed 10-fold cross-validation with 1,000 iterations. We selected Random Forest (RF), Support Vector Machine Radial Kernel (SVMRadial), and Neural Network (NNET) algorithms for test set validation, based on the suitability of our data set, popularity, and reliability. For the RF model, tuning parameters were optimized with bootstrap = TRUE, which resembles random sampling during model building. The maximum number of tree splits in each step was a max_depth = (50, 80, 100, 150, 300), maximum features were selected as auto (max_features = 'auto'), and for error minimization through impurity value (min_impurity_decrease = c (0.0, 0.02, 0.1, 0.5). Next, a minimum tree split as a leaf in each step (min_samples_leaf = (1 to 10), maximum generation of trees (n_estimator = 20), and other parameters as a default. The best fit optimized parameters were considered the final model for further evaluation. For Support Vector Machine (SVM), we tuned two major parameters: 1) cost function, which ensures the decision boundary for data classes, and 2) the sigma value, which defines how much influence a single training set has on the model, with lower sigma and cost resulting in better prediction accuracy. For neural network algorithms, we used hidden layers (size = 1,2,4,6,10,15) and learning rate (decay = 0, 0.05, 0.1, 1, 2) as tuning parameters (17). The prediction statistics and accuracy of RF, SVMRadial and NNET were examined through several statistics such as Area Under the Receiver Operating Characteristic (AUROC), balanced accuracy (kappa), sensitivity, and specificity, which are compiled in **Supplementary Table 3**.

The genotypes of each subject at each of the 200 PSA-associated SNPs identified by Patrick et al. (18) were compiled from imputed subject genotyping data (see section *'Genotype data processing'*). We coded genotypes homozygous for the non-risk allele as zero, heterozygous as one, and homozygous for the risk allele as two. As above, eleven ML algorithms were trained on this data and evaluated based on classification accuracy and kappa, and the performance of three models (RF, SVMRadial, and NNET) were examined through test set data and optimized using the same tuning parameters. The ML algorithms were run with set.seed=862 for reproducibility of models.

RESULTS

Cell Types Enriched and Depleted Among PSA, PSO, and Healthy PBMCs

We characterized the differences in cellular composition as well as transcriptional and cell surface protein expression between 28 PSA, 24 PSO, and 29 healthy subjects, along with 14 psoriasis patients with unclear PSA diagnosis (PSX) by performing single cell RNA-seq on PBMCs, obtaining transcriptomes of 392 – 7003 (median of 2,392) cells per subject (total 246,762 cells). For a subset of these cells (133,665, 54%), we additionally performed antibody-derived tag labeling of 258 cell surface proteins (**Supplementary Table 2**).

We performed integrative mapping of transcriptomic data from our cell population to categorize all cells into 30 phenotypic subsets defined in a previously described multimodal reference dataset of healthy PBMCs (**Figure 1A**) (14). All 30 cell types were comparably represented among PSA, PSO, PSX and healthy subjects (**Figure 1B**), with the exception of Tregs and dnT cells, which were relatively increased in PSA patients compared to both PSO and healthy subjects ($p < 0.03$, **Figure 1B**), and hematopoietic stem precursor cells (HSPCs), which were relatively increased in healthy subjects ($p < 0.007$).

Gene and Protein Biomarkers of PSA Include Activational and Metabolic Transcriptomic Differences That Distinguish PSA From PSO

We next surveyed the phenotypic differences between PSA, PSO, and healthy cells of each cell type by calculating differentially expressed genes (DEGs) and proteins (DEPs). Within each cell subset, we found 1 – 135 DEGs (median 23) and 1 – 18 DEPs (median 4) with significant differences between PSA and PSO, PSA and healthy, or PSO and healthy cells (**Figures 2A, B**), with the most differentially expressed features detected in CD14 monocytes, the most abundant cell type in our dataset.

The DEGs and DEPs represented both broad as well as cell type-specific disease-associated expression differences. Among 30 DEGs with the highest absolute fold change (**Figure 2C**), we observed a general upregulation of mitochondrial genes (*MT-CO3, MT-ND1*, *MT-ND3*) paired with a downregulation of ribosomal protein gene *RPS26* across most cell types in PSA patients relative to PSO patients or healthy subjects. Among PSA T and NK cells, we also observed a downregulation of AP-1 transcription factors (*JUN, JUNB, JUND, FOS*) and regulators of activation (*TNFAIP3, DUSP1*) along with the upregulation of *S100A11*, a calcium-binding protein associated with rheumatoid arthritis (19). Lastly, we observed PSA-associated differences in chemokine receptor expression, specifically a downregulation of *CXCR4* in T and NK cells and an upregulation of *CX3CR1* in monocytes, NK cells, and specific T cell subsets. Disease-associated differences in cell surface protein expression, in contrast, were more sparsely observed within specific cell types (**Figure 2D**). Among the top 30 DEPs, HLA-A2 was broadly upregulated among B and T cell subsets as well as CD14 monocytes in PSA patients, while CD205 was

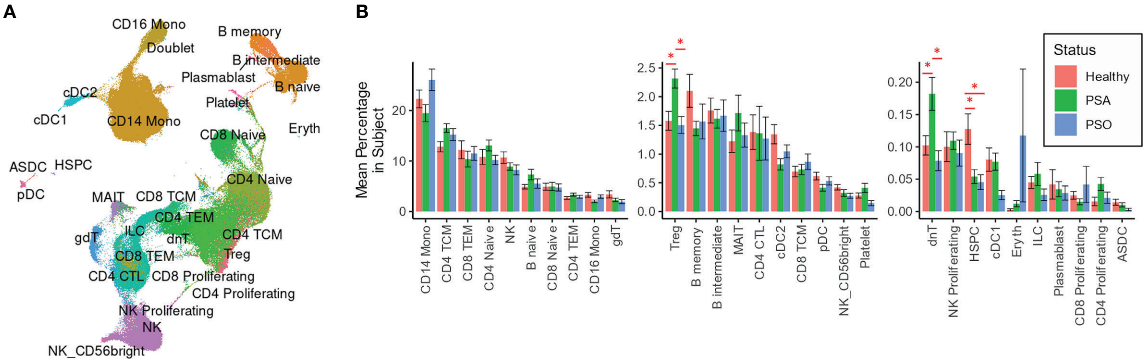


FIGURE 1 | Cell types and subsets among PSA, PSO, and healthy individuals. **(A)** UMAP of SCTransform-normalized RNA expression integrated with ADT expression, colored by cell subset. **(B)** Mean percentages of each cell type within the total PBMCs of each subject. Error bars indicate standard error of the mean; * indicates both Wilcoxon and FDR-adjusted Kruskall-Wallis p-values < 0.05.

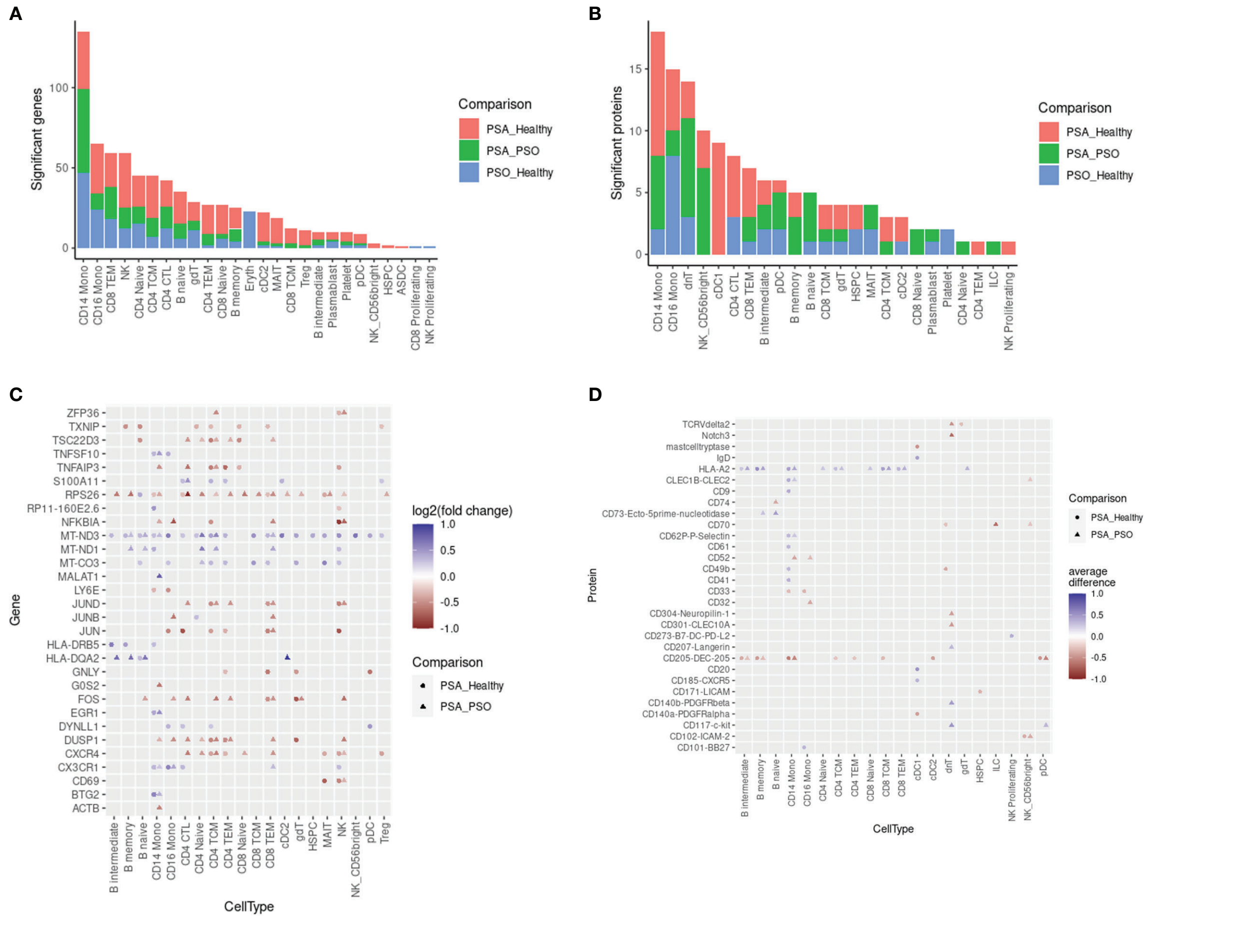


FIGURE 2 | Differentially expressed features between PSA, PSO, and healthy subjects within cell types. Counts of differentially expressed **(A)** genes and **(B)** cell surface proteins are shown for each comparison within each cell type. Top 30 differentially expressed **(C)** genes and **(D)** cell surface proteins in each cell type are ranked by highest absolute log2 fold change (for genes) or absolute mean difference (for proteins) between PSA cells vs. PSO (circles) or healthy (triangles) cells.

broadly downregulated in many of the same cell types along with cDC2 and pDC subsets.

Phenotypic Subsets of Specific Cell Types Enriched and Depleted in PSA

Besides differences in cellular composition and gene or protein expression, we also searched for additional disease signatures of PSA among the phenotypic subsets of each cell type. By performing integrative, transcriptome-based *de novo* clustering of each cell type with at least 1,000 cells, we identified phenotypic clusters within six cell types that were enriched in different disease conditions (**Figure 3A**).

Some of these phenotypic subsets were uniquely associated with PSA. Within CD16 monocytes, a small cluster (cluster 7) was more abundant among PSA subjects (**Figure 3B**). Compared to other CD16 monocytes, the 108 cells of this cluster showed generally lower expression of several mitochondrial genes (**Figure 3C** and **Supplementary Table 5**) and higher expression of S100 genes (*S100A4, S100A6, S100A10, S100A11*), as well as genes involved in antigen presentation (*HLA-DRB5*, *HLA-DQB1*, *FCER1G*) and regulation of innate activation [*DUSP1* (20)]. We also observed a cluster of PSA-abundant MAIT cells (cluster 2), however, these cells may potentially represent a clustering artifact, as no significantly over- or under-expressed genes were found to distinguish this cluster from other cells. Analysis of differentially expressed proteins in these two clusters yielded a single protein, Tetraspanin 33, which was under-expressed in CD16 monocyte cluster 7.

On the other hand, we also found clusters uniquely reduced in PSA within B memory (cluster 1) and a CD4 TEM (cluster 2) cells (**Figure 3B**). The B memory cluster was characterized by a small number of gene expression differences including reduced expression of *IGLC2* and *IGLC3* that was consistent with downregulated cell surface expression of immunoglobin light chain protein (**Figure 3C** and **Supplementary Table 5**). Additionally, we observed increased expression of *JUNB*, a negative regulator of growth and proliferation (21) and downregulated expression of transferrin receptor [CD71, a marker of activated or proliferating B cells (22)] and several other receptors found to promote apoptosis and proliferation [CD95 (23), CD164 (24)] or response to chemokines (CD99 (25)].

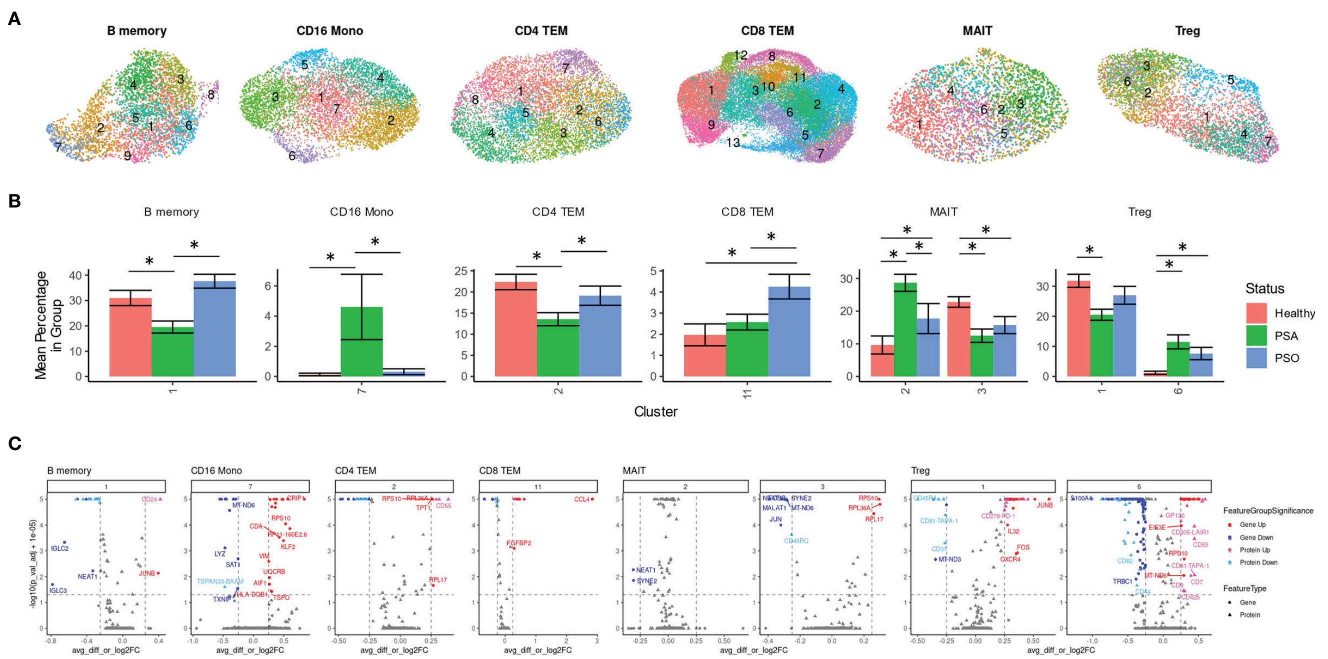


FIGURE 3 | Immune cell subsets differentially abundant in psoriatic and healthy individuals. **(A)** UMAP of *de novo* clusters identified within select cell types containing clusters with significant abundance differences. **(B)** Average percentage of cells from each PSA, PSO, or healthy subject in a given cluster out of total cells from that subject in the given cell type. **(C)** Volcano plots of genes and cell surface proteins upregulated and downregulated in each cluster relative to other cells of the same cell type. * indicates Wilcoxon p-value < 0.05.

The CD4 TEM cluster showed a downregulation of *DUSP2*, a negative regulator of Th17 differentiation (26), as well as Jun/Fos genes (*JUN, JUNB, FOS, FOSB*) and, unexpectedly, several genes associated with cytotoxic function (*GZMA, GZMK, NKG7, SRGN*) (**Supplementary Table 5**). Differential protein analysis revealed an upregulation of gut-homing integrin β7 and receptors that promote cell proliferation [CD55 (27)] and maintain T cell survival [CD127 (28)].

Other clusters were associated specifically with PSO or healthy subjects. Cells within a single CD8 TEM cluster enriched among PSO subjects (cluster 11, **Figure 3B**) showed a strong upregulation of *CCL4*, a CD8$^+$ T cell recruiting (29) chemokine associated with psoriasis (30), along with other inflammatory cytokines and chemokines (*TNF, IFNG, CCL3, CCL4L2*) (**Figure 3C** and **Supplementary Table 5**). Differential expression analysis of cell surface proteins on this cluster revealed an upregulation of GPR56, a marker of cytotoxic cells (31) as well as reduced expression of chemokine receptor CXCR3. We also found a MAIT cluster (cluster 3) enriched among healthy subjects, though, similar to MAIT cluster 2 above, these cells are distinguished by relatively few markers that included ribosomal proteins and long non-coding RNAs *NEAT1* and *MALAT1* (**Figure 3C** and **Supplementary Table 5**).

Lastly, clustering analysis among Tregs revealed an imbalance of resting and activated Tregs between healthy and psoriatic (PSO and PSA) subjects. Differentially expressed genes in a Treg cluster enriched in PSA (cluster 6, **Figure 3B**) consisted of an upregulation of 52 genes that mostly encoded ribosomal proteins and a downregulation of 115 genes, including some involved in class I and class II antigen presentation (*HLA-A, HLA-B, HLA-C, HLA-E, HLA-DPA1, HLA-DPB1, HLA-DRA, HLA-DRB1*) and *CD52* (**Supplementary Table 5**), which encodes a costimulatory receptor found to promote Treg suppression of CD4 and CD8 T cells (32). Differential expression of cell surface proteins also revealed a lower expression of memory marker CD45RO, which, combined with higher expression of CD45RA and *CCR7* in this cluster (**Figure 3C** and **Supplementary Table 5**), suggests a naïve, or antigen-inexperienced state. We additionally observed an upregulation of GP130 (**Figure 3B**), a subunit of multiple cytokine receptors such as IL-6R that has been found to define a Treg subset with reduced suppression function (33). These protein expression differences were reversed in the relatively healthy-enriched cluster 1, in which GP130 and CD45RA were reduced in expression while CD45RO, along with costimulatory markers such as TIGIT and PD-1, were increased in expression. DEGs from this cluster, including an upregulation of *DUSP1*, *CXCR4* and Jun and Fos family genes (**Figure 3C** and **Supplementary Table 5**) further suggested an activated, functionally suppressive phenotype, and *FOXP3* expression was higher (though not significantly) in this cluster than cluster 6 (**Supplementary Table 5**).

Machine Learning Classifiers Distinguish Between PSA, PSO, and Healthy Subjects Using Cell Type-Specific DEGs and DEPs

We evaluated the diagnostic potential of the PSA-associated DEGs and DEPs by using them to perform ML classification of subjects in our study cohort. Based on the cell-type specific mean expression of 257 DEGs and 258 DEPs (**Supplementary Table 4**) averaged within each subject's cells in the corresponding cell types, we performed ensemble feature selection (16) using four ML algorithms to identify the top twenty DEGs and DEPs based on their classification rate (Importance).

The top twenty DEGs span a variety of immune cell types, and many encode proteins involved in metabolism, translation, and transcriptional regulation (**Figure 4A**). 10-fold cross validation of eleven ML algorithms trained on these features classified PSA, PSO, and healthy subjects with average accuracies of 0.65 – 0.89 (**Figure 4B**) across 1,000 iterations. Kappa, a measure of the agreement between observed and expected accuracy ranged from 0.41- 0.72 across the eleven algorithms. Further evaluation of the sensitivity and specificity of the RF model demonstrated an AUROC of 0.89, 0.99, and 0.87 for healthy, PSA, and PSO subjects (**Figure 4C**) with similar results for SVMRadial and NNET models (**Supplementary Figure 4**).

Similar to the top twenty DEGs, the top twenty DEPs spanned several cell types but showed generally lower classification Importance (**Figure 4D**). Accordingly, the average accuracy of the eleven ML algorithms on the top twenty DEPs (0.58 - 0.86, **Figure 4E**) and kappa (0.21 - 0.65) were relatively lower than for DEGs, with similarly reduced AUROC for RF (Healthy = 0.80, PSA = 0.93, PSO = 0.88, **Figure 4F**), SVMRadial, and NNET (**Supplementary Figure 4**).

To test whether classifier performance could be improved by considering both gene and cell surface protein expression together, we also performed ensemble feature selection on the combined expression data of DEPs and DEGs from the above analyses. The resulting set of twenty features consisted of 10 DEGs and 10 DEPs spanning several cell types, with similar

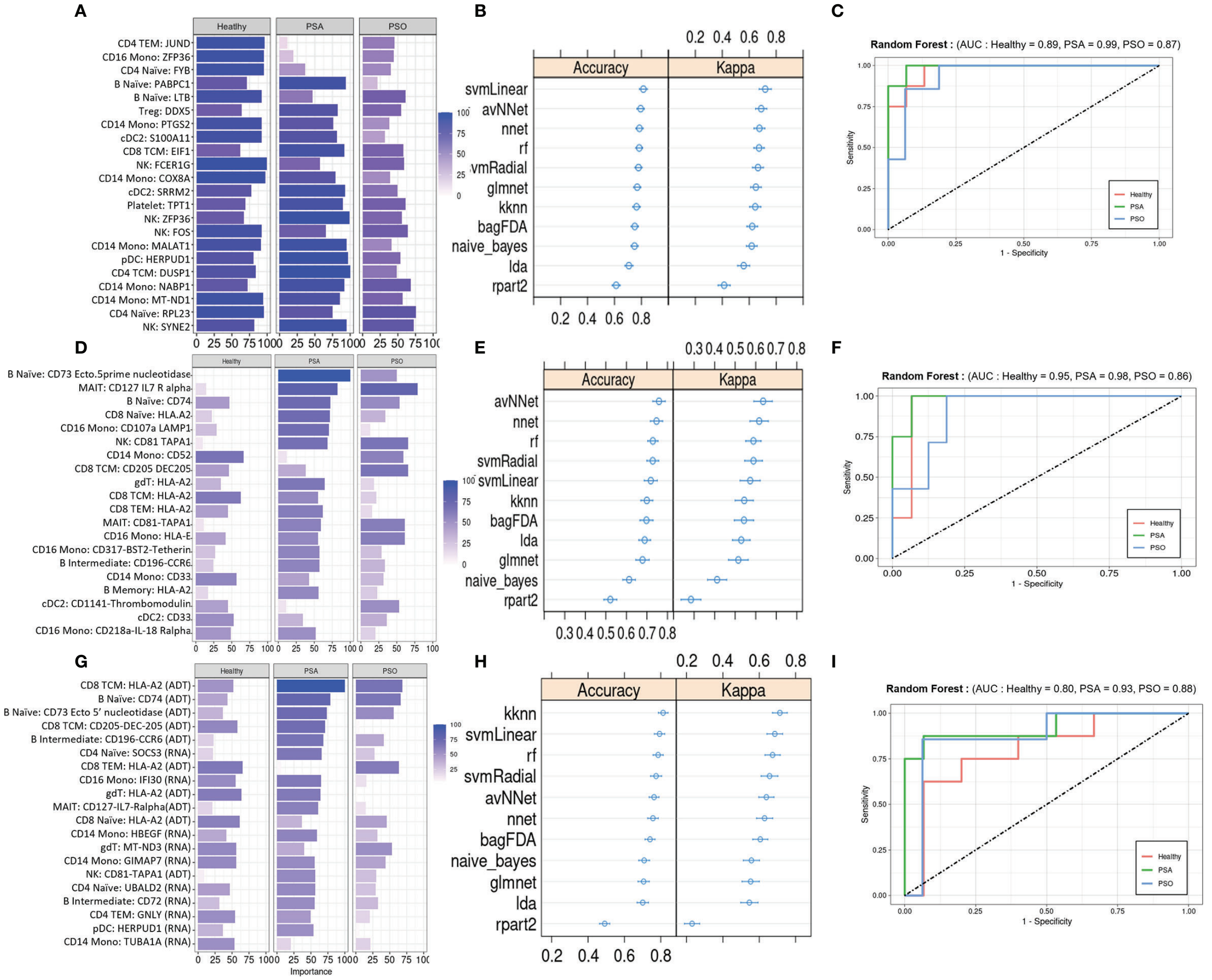


FIGURE 4 | Machine learning classification of healthy, PSA and PSO subjects. **(A)** Classification rate (Importance) of top 20 DEGs, along with corresponding **(B)** accuracy and kappa of eleven different ML classifiers trained on these features. **(C)** ROC curve of RF model for healthy, PSA, and PSO classification. Analogous plots shown for **(D–F)** DEPs and **(G–I)** DEGs combined with DEPs. Error bars indicate 95% confidence interval.

classification Importance measures as the set of DEPs only (**Figure 4G**). Classification accuracies of the eleven ML algorithms based on this feature set were more comparable to those of DEGs alone (accuracy 0.68 - 0.89, kappa 0.54 – 0.76, **Figure 4H**), except for rpart which performed worse on this feature set than on DEGs and DEPs separately (average accuracy 0.52, kappa 0.26). AUROC for RF was relatively lower than DEP- and DEG-only models for healthy (0.80) and PSA (0.93) groups but similar to those models for PSO (0.88) subjects (**Figure 4I**), with comparable results for SVMRadial and NNET models (**Supplementary Figure 4**).

We performed further validation of the RF model by using it to classify a cohort of 14 subjects (PSX) presenting with cutaneous psoriasis and joint pain that did not confidently meet current PSA diagnosis criteria. RF classification based on DEGs, DEPs, or both consistently categorized 10 of these patients as PSA and one patient as healthy (**Supplementary Figure 5** and **Supplementary Table 6**), with the remaining three subjects discordantly classified as healthy or PSO.

ML Classifiers Detect PSA in Psoriatic Individuals Using DEGs, DEPs, or Genetic Risk Factors

We also evaluated the diagnostic potential of DEGs and DEPs, separately or in combination, for detecting PSA among individuals presenting with cutaneous psoriasis by performing a two-way classification of PSA and PSO groups. As before, the top twenty DEGs and DEPs were associated with several immune cell types, with the DEG set including many genes with roles in metabolism and in the regulation of activation and inflammation (**Figure 5A**). Among PSA and PSO subjects, we noted higher Importance measures among the top twenty DEPs compared with the top twenty DEGs (**Figures 5A, D**), however performance metrics of the eleven ML models were generally higher in DEGs (accuracy 0.81 – 0.94, kappa 0.41 – 0.83) than DEPs (accuracy 0.73 – 0.92, kappa 0.42 – 0.72, **Figures 5B, E**). In addition, RF, SVMRadial, and NNET all achieved perfect classification of PSA and PSO subjects using DEGs (AUROC of 1, **Figure 5C** and **Supplementary Figures 6A, B**) compared to the slightly lower classification performance for DEPs (**Figure 5F** and **Supplementary Figures 6C, D**). Feature selection on combined DEPs and DEGs yielded a top twenty feature set with Importance measures that were intermediate between the sets of DEPs and DEGs alone (**Figure 5G**), and while ML classifier performance was lower for the combined feature set (accuracy 0.52 – 0.81, kappa 0.26 – 0.67, **Figure 5H**), AUROC for the RF and SVMRadial models (1.00 and 0.96, respectively, **Figure 5I** and **Supplementary Figure 6E**) was comparable to those of DEG- and DEP-only feature sets, with NNET underperforming substantially (AUROC 0.7, **Supplementary Figure 6F**).

Lastly, we evaluated whether our ML framework for detecting PSA in a background of cutaneous psoriasis could also be applied to genetic biomarkers of PSA risk. ML classifiers trained on patient genotypes at 200 SNP sites previously found to be associated with PSA (18) achieved average classification accuracies between 0.6 and 0.87 and kappa between 0.51 and 0.73 (**Figure 6A**). AUROC of RF, SVM-Radial, and NNET was 0.92 (**Figure 6B**), 0.88, and 0.81, similar to metrics reported in the previous study (18).

DISCUSSION

Our study sheds light on the phenotypic differences between the circulating immune cells of PSA and PSO patients at multiple levels of resolution. At the cellular level, we observed a higher abundance of Tregs and dnT cells in PSA patients and a higher abundance of HSPCs in healthy subjects. While, to our knowledge, the role of dnTs and HSPCs in PSA has not been extensively investigated, dnT cells have been reported to infiltrate psoriatic skin as well as participate in IL-23/IL-17 signaling in mouse models of psoriasis (34) and spondyloarthritis (35), and the proliferation and differentiation of HSPCs is currently known to respond to systemic interferon and TNF signaling (36). While we observed increased peripheral Tregs in PSA patients, whether this subset is generally increased or decreased in PSA is still unclear in light of conflicting results found in other studies (37, 38).

Within each cell type, our *de novo* clustering analyses identified disease-associated subsets and potential biological processes affecting them. First, the skewing of peripheral Tregs toward more naïve, resting cells and fewer activated effector cells in PSA and PSO parallels what has been observed in systemic lupus erythematosus (39) and could reflect either a migration of effector Tregs from circulation into sites of inflammation or a general expansion of the naïve Treg pool. Second, our study also identified a cluster of CD8 TEM cells specific to PSO but not PSA. The strong upregulation of *CCL4* coupled with the downregulation of CXCR3 in this cluster raise the possibility that differences in chemokine-mediated immune cell homing (e.g. to skin compared with synovium) could emerge as a key characteristic for predicting PSA progression or risk in PSO patients, especially in light of evidence suggesting that CXCR3 may be involved in T cell recruitment in PSA, based on higher protein expression of its ligands, CXCL9 and CXCL10, in synovial compared with peripheral compartments (40, 41), whereas no such difference was found for MIP1β, the chemokine encoded by *CCL4* (41). Besides the overall PSA-associated downregulation of *CXCR4* and upregulation of *CX3CR1* observed in our data, other studies have identified PSO- and PSA-associated T cell subsets expressing CCR5 (42), CCR4 (43), and CCR10 (37), and the questions of whether and how signaling through these chemokine receptors mediates trafficking of pathological T cells between the skin, blood, and joint remain active areas of investigation. Lastly, the enrichment of a CD16 monocyte subset that we observed in PSA subjects is consistent with previous findings of increased circulating CD16+ monocyte population in PSA subjects that can give rise to osteoclasts (44). Other studies in mice have found that subsets of other myeloid cell types, such as neutrophils, may also contribute to psoriatic disease through T-cell independent responses to IL-17A signaling (45, 46).

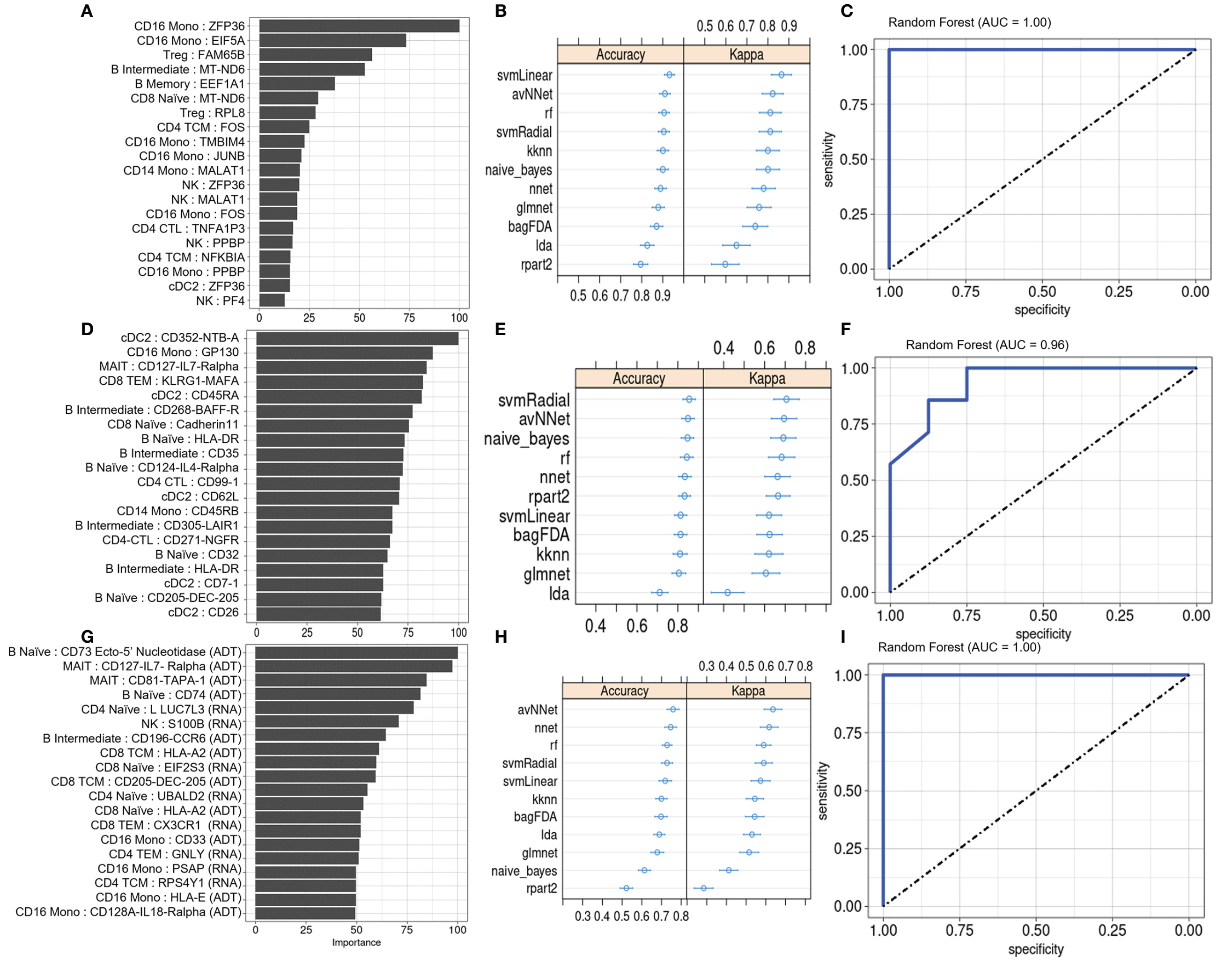


FIGURE 5 | Machine learning classification of PSA vs. PSO subjects. **(A)** Classification rate (Importance) of top 20 DEGs, along with corresponding **(B)** accuracy and kappa of eleven different ML classifiers trained on these features. **(C)** ROC curve of RF model. Analogous plots shown for **(D–F)** DEPs and **(G–I)** DEGs combined with DEPs. Error bars indicate 95% confidence interval.

At the molecular level, we found disease-associated protein and gene expression signatures within diverse innate and adaptive immune cell types, consistent with the current understanding that multiple cell types contribute to inflammation in PSA (6). While these contributions have mainly been investigated in the context of IL-17 and IL-23 signaling, our data sheds light on other characteristics that distinguish circulating immunocytes in PSA patients, such as generally increased mitochondrial gene expression and decreased expression of cell activational regulators. Although disease conditions may generally alter protein and gene expression divergently among different cell types, we note that, in our data, gene and protein expression in different cell types are largely perturbed in the same direction by PSA (i.e. a feature upregulated in PSA cells of one type is generally upregulated in PSA cells of other types).

Importantly, our study demonstrates the utility of cell-specific gene and cell surface protein expression differences when incorporated into a ML framework for detecting PSA, with most of the ML algorithms considered in this study classifying PSA, PSO, and healthy subjects or distinguishing just between PSA and PSO subjects with >70% average accuracy on either gene or protein features. Combining both types of features reduced overall model performance, possibly due to differences in the magnitude of interindividual or technical variation in the detected expression of these feature types, which may not be accurately accounted for by the subject-averaged expression data we used for model training and testing. Nevertheless, our study expands the number of potential biomarkers and cell types relevant to diagnosing PSA and understanding its biology.

We note that our data, being derived solely from peripheral blood immune populations, cannot address whether these cell

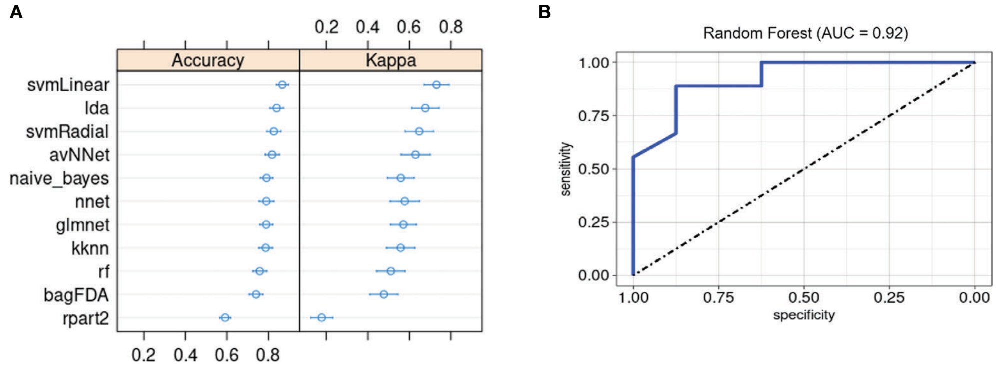


FIGURE 6 | Machine learning classification of PSA vs. PSO subjects based on 200 PSA-associated genetic risk loci. **(A)** Accuracy and kappa of eleven ML models. **(B)** ROC curve for RF model. Error bars indicate 95% confidence interval. Error bars indicate 95% confidence interval.

types are also present in the synovium, whether they may instead represent systemic responses to cutaneous inflammation in PSA subjects, or the extent that they arise from either a migration of cells between blood and tissue compartments or an overall expansion or reduction in specific cell subsets. Also, since PSA patients in our study already have established arthritic disease, our data may not capture early or ephemeral biomarkers of disease that may appear in PSO patients who eventually develop PSA. Future investigations combining single cell multiomics on blood, skin, and joint immune populations with a longitudinal follow-up of PSO patients [as employed by Abji et al. (47)] may help overcome these limitations.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: NCBI GEO, accession no: GSE194315 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE194315).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of California San Francisco IRB. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

WL and LG conceived and supervised the project. JH, DP, MCas, MCal, H-WC, MCh, SY, EB, MH, TB, MM, LG, and WL recruited study subjects and performed clinical annotation. H-WC and Z-MH performed experimental procedures. JL, SK, DC, and WL performed data analysis. CY provided technical and analytic support. JL, SK, and WL wrote and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by grants from the National Psoriasis Foundation to WL. This study was also supported by PREMIER, a NIH/NIAMS P30 Center for the Advancement of Precision Medicine in Rheumatology at UCSF (P30AR070155). WL is supported by NIH grant R01AR078688. JL is supported by NIH grant T32AR007175. CY is supported by NIH grants R01AR071522, R01AI136972, U01HG012192, and the Chan Zuckerberg Initiative, and is an investigator at the Chan Zuckerberg Biohub and is a member of the Parker Institute for Cancer Immunotherapy (PICI). The funders above were not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

ACKNOWLEDGMENTS

We would like to thank Laura Trupin, Jinoos Yazdany, and Mary Nakamura for their administrative support as well as Yang Sun and George Hartoularos for technical guidance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.835760/full#supplementary-material

REFERENCES

1. Mease PJ, Gladman DD, Papp KA, Khraishi MM, Thaçi D, Behrens F, et al. Prevalence of Rheumatologist-Diagnosed Psoriatic Arthritis in Patients With Psoriasis in European/North American Dermatology Clinics. *J Am Acad Dermatol* (2013) 69:729–35. doi: 10.1016/j.jaad.2013.07.023
2. Haroon M, Gallagher P, FitzGerald O. Diagnostic Delay of More Than 6 Months Contributes to Poor Radiographic and Functional Outcome in Psoriatic Arthritis. *Ann Rheumatic Dis* (2015) 74:1045–50. doi: 10.1136/annrheumdis-2013-204858
3. Pennington SR, FitzGerald O. Early Origins of Psoriatic Arthritis: Clinical, Genetic and Molecular Biomarkers of Progression From Psoriasis to Psoriatic Arthritis. *Front Med* (2021) 8:723944. doi: 10.3389/fmed.2021.723944
4. Stuart PE, Nair RP, Tsoi LC, Tejasvi T, Das S, Kang HM, et al. Genome-Wide Association Analysis of Psoriatic Arthritis and Cutaneous Psoriasis Reveals Differences in Their Genetic Architecture. *Am J Hum Genet* (2015) 97:816–36. doi: 10.1016/j.ajhg.2015.10.019
5. FitzGerald O, Haroon M, Giles JT, Winchester R. Concepts of Pathogenesis in Psoriatic Arthritis: Genotype Determines Clinical Phenotype. *Arthritis Res Ther* (2015) 17:115. doi: 10.1186/s13075-015-0640-3
6. O'Brien-Gore C, Gray EH, Durham LE, Taams LS, Kirkham BW. Drivers of Inflammation in Psoriatic Arthritis: The Old and the New. *Curr Rheumatol Rep* (2021) 23:40. doi: 10.1007/s11926-021-01005-x
7. Cretu D, Gao L, Liang K, Soosaipillai A, Diamandis EP, Chandran V. Differentiating Psoriatic Arthritis From Psoriasis Without Psoriatic Arthritis Using Novel Serum Biomarkers. *Arthritis Care Res* (2018) 70:454–61. doi: 10.1002/acr.23298
8. Chandran V, Cook RJ, Edwin J, Shen H, Pellett FJ, Shanmugarajah S, et al. Soluble Biomarkers Differentiate Patients With Psoriatic Arthritis From Those With Psoriasis Without Arthritis. *Rheumatology* (2010) 49:1399–405. doi: 10.1093/rheumatology/keq105
9. Leijten E, Tao W, Pouw J, van Kempen T, Olde Nordkamp M, Balak D, et al. Broad Proteomic Screen Reveals Shared Serum Proteomic Signature in Patients With Psoriatic Arthritis and Psoriasis Without Arthritis. *Rheumatology* (2021) 60:751–61. doi: 10.1093/rheumatology/keaa405
10. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* (2007) 81:559–75. doi: 10.1086/519795
11. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets. *GigaScience* (2015) 4:7. doi: 10.1186/s13742-015-0047-8
12. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The Variant Call Format and VCFtools. *Bioinformatics* (2011) 27:2156–8. doi: 10.1093/bioinformatics/btr330
13. Kang HM, Subramaniam M, Targ S, Nguyen M, Maliskova L, McCarthy E, et al. Multiplexed Droplet Single-Cell RNA-Sequencing Using Natural Genetic Variation. *Nat Biotechnol* (2018) 36:89–94. doi: 10.1038/nbt.4042
14. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated Analysis of Multimodal Single-Cell Data. *Cell* (2021) 184:3573–3587.e29. doi: 10.1016/j.cell.2021.04.048
15. DePasquale EAK, Schnell DJ, Van Camp P-J, Valiente-Alandí Í, Blaxall BC, Grimes HL, et al. DoubletDecon: Deconvoluting Doublets From Single-Cell RNA-Sequencing Data. *Cell Rep* (2019) 29:1718–27.e8. doi: 10.1016/j.celrep.2019.09.082
16. Hoque N, Singh M, Bhattacharyya DK. EFS-MI: An Ensemble Feature Selection Method for Classification. *Complex Intelligent Syst* (2018) 4:105–18. doi: 10.1007/s40747-017-0060-x
17. Yoo Y. Hyperparameter Optimization of Deep Neural Network Using Univariate Dynamic Encoding Algorithm for Searches. *Knowledge-Based Syst* (2019) 178:74–83. doi: 10.1016/j.knosys.2019.04.019
18. Patrick MT, Stuart PE, Raja K, Gudjonsson JE, Tejasvi T, Yang J, et al. Genetic Signature to Provide Robust Risk Assessment of Psoriatic Arthritis Development in Psoriasis Patients. *Nat Commun* (2018) 9:4178. doi: 10.1038/s41467-018-06672-6
19. Andrés Cerezo L, Šumová B, Prajzlerová K, Veigl D, Damgaard D, Nielsen CH, et al. Calgizzarin (S100A11): A Novel Inflammatory Mediator Associated With Disease Activity of Rheumatoid Arthritis. *Arthritis Res Ther* (2017) 19:79. doi: 10.1186/s13075-017-1288-y
20. Abraham SM, Clark AR. Dual-Specificity Phosphatase 1: A Critical Regulator of Innate Immune Responses. *Biochem Soc Trans* (2006) 34:1018–23. doi: 10.1042/BST0341018
21. Szremska AP, Kenner L, Weisz E, Ott RG, Passegué E, Artwohl M, et al. JunB Inhibits Proliferation and Transformation in B-Lymphoid Cells. *Blood* (2003) 102:4159–65. doi: 10.1182/blood-2003-03-0915
22. Ellebedy AH, Jackson KJL, Kissick HT, Nakaya HI, Davis CW, Roskin KM, et al. Defining Antigen-Specific Plasmablast and Memory B Cell Subsets in Human Blood After Viral Infection or Vaccination. *Nat Immunol* (2016) 17:1226–34. doi: 10.1038/ni.3533
23. Koncz G, Hueber A-O. The Fas/CD95 Receptor Regulates the Death of Autoreactive B Cells and the Selection of Antigen-Specific B Cells. *Front Immunol* (2012) 3:207. doi: 10.3389/fimmu.2012.00207
24. Tu M, Cai L, Zheng W, Su Z, Chen Y, Qi S. CD164 Regulates Proliferation and Apoptosis by Targeting PTEN in Human Glioma. *Mol Med Rep* (2017) 15:1713–21. doi: 10.3892/mmr.2017.6204
25. Gil M, Pak H-K, Lee A-N, Park S-J, Lee Y, Roh J, et al. CD99 Regulates CXCL12-Induced Chemotaxis of Human Plasma Cells. *Immunol Lett* (2015) 168:329–36. doi: 10.1016/j.imlet.2015.10.015
26. Lu D, Liu L, Ji X, Gao Y, Chen X, Liu Y, et al. The Phosphatase DUSP2 Controls the Activity of the Transcription Activator STAT3 and Regulates TH17 Differentiation. *Nat Immunol* (2015) 16:1263–73. doi: 10.1038/ni.3278
27. Capasso M, Durrant LG, Stacey M, Gordon S, Ramage J, Spendlove I. Costimulation *via* CD55 on Human CD4+ T Cells Mediated by CD97. *J Immunol (Baltimore Md : 1950)* (2006) 177:1070–7. doi: 10.4049/jimmunol.177.2.1070
28. Miller ML, Mashayekhi M, Chen L, Zhou P, Liu X, Michelotti M, et al. Basal NF- B Controls IL-7 Responsiveness of Quiescent Naive T Cells. *Proc Natl Acad Sci* (2014) 111:7397–402. doi: 10.1073/pnas.1315398111
29. Castellino F, Huang AY, Altan-Bonnet G, Stoll S, Scheinecker C, Germain RN. Chemokines Enhance Immunity by Guiding Naive CD8+ T Cells to Sites of CD4+ T Cell–Dendritic Cell Interaction. *Nature* (2006) 440:890–5. doi: 10.1038/nature04651
30. Pedrosa E, Carretero-Iglesia L, Boada A, Colobran R, Faner R, Pujol-Autonell I, et al. CCL4L Polymorphisms and CCL4/CCL4L Serum Levels Are Associated With Psoriasis Severity. *J Invest Dermatol* (2011) 131:1830–7. doi: 10.1038/jid.2011.127
31. Peng Y-M, van de Garde MDB, Cheng K-F, Baars PA, Remmerswaal EBM, van Lier RAW, et al. Specific Expression of GPR56 by Human Cytotoxic Lymphocytes. *J leukocyte Biol* (2011) 90:735–40. doi: 10.1189/jlb.0211092
32. Watanabe T, Masuyama J, Sohma Y, Inazawa H, Horie K, Kojima K, et al. CD52 is a Novel Costimulatory Molecule for Induction of CD4+ Regulatory T Cells. *Clin Immunol* (2006) 120:247–59. doi: 10.1016/j.clim.2006.05.006
33. Dhuban KB, Bartolucci S, D'Hennezel E, Piccirillo CA. Signaling Through Gp130 Compromises Suppressive Function in Human FOXP3+Regulatory T Cells. *Front Immunol* (2019) 10:1532. doi: 10.3389/fimmu.2019.01532
34. Ueyama A, Imura C, Fusamae Y, Tsujii K, Furue Y, Aoki M, et al. Potential Role of IL-17-Producing CD4/CD8 Double Negative αβ T Cells in Psoriatic Skin Inflammation in a TPA-Induced STAT3C Transgenic Mouse Model. *J Dermatol Sci* (2017) 85:27–35. doi: 10.1016/j.jdermsci.2016.10.007
35. Sherlock JP, Joyce-Shaikh B, Turner SP, Chao C-C, Sathe M, Grein J, et al. IL-23 Induces Spondyloarthropathy by Acting on ROR-γt+ CD3+CD4-CD8- Entheseal Resident T Cells. *Nat Med* (2012) 18:1069–76. doi: 10.1038/nm.2817
36. King KY, Goodell MA. Inflammatory Modulation of HSCs: Viewing the HSC as a Foundation for the Immune Response. *Nat Rev Immunol* (2011) 11:685–92. doi: 10.1038/nri3062
37. Leijten EF, van Kempen TS, Olde Nordkamp MA, Pouw JN, Kleinrensink NJ, Vincken NL, et al. Tissue-Resident Memory CD8+ T Cells From Skin Differentiate Psoriatic Arthritis From Psoriasis. *Arthritis Rheumatol* (2021) 73:1220–32. doi: 10.1002/art.41652
38. Wang J, Zhang S-X, Hao Y-F, Qiu M-T, Luo J, Li Y-Y, et al. The Numbers of Peripheral Regulatory T Cells are Reduced in Patients With Psoriatic Arthritis and Are Restored by Low-Dose Interleukin-2. *Ther Adv Chronic Dis* (2020) 11:204062232091601. doi: 10.1177/2040622320916014
39. Pan X, Yuan X, Zheng Y, Wang W, Shan J, Lin F, et al. Increased CD45RA+ FoxP3low Regulatory T Cells With Impaired Suppressive Function in Patients

With Systemic Lupus Erythematosus. *PloS One* (2012) 7:e34662. doi: 10.1371/journal.pone.0034662

40. Diani M, Casciano F, Marongiu L, Longhi M, Altomare A, Pigatto PD, et al. Increased Frequency of Activated CD8+ T Cell Effectors in Patients With Psoriatic Arthritis. *Sci Rep* (2019) 9:1–10. doi: 10.1038/s41598-019-47310-5
41. Penkava F, Velasco-Herrera MDC, Young MD, Yager N, Nwosu LN, Pratt AG, et al. Single-Cell Sequencing Reveals Clonal Expansions of Pro-Inflammatory Synovial CD8 T Cells Expressing Tissue-Homing Receptors in Psoriatic Arthritis. *Nat Commun* (2020) 11:4767. doi: 10.1038/s41467-020-18513-6
42. Sgambelluri F, Diani M, Altomare A, Frigerio E, Drago L, Granucci F, et al. A Role for CCR5(+)CD4 T Cells in Cutaneous Psoriasis and for CD103(+) CCR4(+) CD8 Teff Cells in the Associated Systemic Inflammation. *J Autoimmun* (2016) 70:80–90. doi: 10.1016/j.jaut.2016.03.019
43. Casciano F, Diani M, Altomare A, Granucci F, Secchiero P, Banfi G, et al. CCR4+ Skin-Tropic Phenotype as a Feature of Central Memory CD8+ T Cells in Healthy Subjects and Psoriasis Patients. *Front Immunol* (2020) 11:529. doi: 10.3389/fimmu.2020.00529
44. Chiu YG, Shao T, Feng C, Mensah KA, Thullen M, Schwarz EM, et al. CD16 (Fcrγiii) as a Potential Marker of Osteoclast Precursors in Psoriatic Arthritis. *Arthritis Res Ther* (2010) 12:R14. doi: 10.1186/ar2915
45. Suzuki E, Maverakis E, Sarin R, Bouchareychas L, Kuchroo VK, Nestle FO. Adamopoulos IE. T Cell-Independent Mechanisms Associated With Neutrophil Extracellular Trap Formation and Selective Autophagy in IL-17a-Mediated Epidermal Hyperplasia. *J Immunol (Baltimore Md : 1950)* (2016) 197:4403–12. doi: 10.4049/jimmunol.1600383
46. Adamopoulos IE, Suzuki E, Chao C-C, Gorman D, Adda S, Maverakis E, et al. IL-17A Gene Transfer Induces Bone Loss and Epidermal Hyperplasia Associated With Psoriatic Arthritis. *Ann Rheumatic Dis* (2015) 74:1284–92. doi: 10.1136/annrheumdis-2013-204782
47. Abji F, Pollock RA, Liang K, Chandran V, Gladman DD. Brief Report: CXCL10 Is a Possible Biomarker for the Development of Psoriatic Arthritis Among Patients With Psoriasis. *Arthritis Rheumatol* (2016) 68:2911–6. doi: 10.1002/art.39800

Conflict of Interest: Author WL has received funding from Abbvie, Amgen, Janssen, Leo, Novartis, Pfizer, Regeneron, and TRex Bio. Author CY is a Scientific Advisory Board member for and holds equity in Related Sciences and ImmunAI, a consultant for and holds equity in Maze Therapeutics, and a consultant for TReX Bio. Author CY has received funding from Chan Zuckerberg Initiative, Chan Zuckerberg Biohub, and Genentech. Author LG has received funding from Novartis, Pfizer, UCB, AbbVie, Eli Lilly, Gilead, Janssen, Novartis, Pfizer and UCB.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

ORIGINAL RESEARCH
published: 15 March 2022
doi: 10.3389/fimmu.2022.784463

Insight Into the Long Noncoding RNA and mRNA Coexpression Profile in the Human Blood Transcriptome Upon *Leishmania infantum* Infection

Sandra Regina Maruyama[1*], *Carlos Alessandro Fuzo*[2], *Antonio Edson R. Oliveira*[3], *Luana Aparecida Rogerio*[1], *Nayore Tamie Takamiya*[1], *Gabriela Pessenda*[4], *Enaldo Vieira de Melo*[5], *Angela Maria da Silva*[5], *Amélia Ribeiro Jesus*[5], *Vanessa Carregaro*[4], *Helder I. Nakaya*[6], *Roque Pacheco Almeida*[5†] and *João Santana da Silva*[4,7†]

[1] Department of Genetics and Evolution, Center for Biological Sciences and Health, Federal University of São Carlos, São Carlos, Brazil, [2] Department of Clinical Analyses, Toxicology and Food Sciences, Ribeirão Preto School of Pharmaceutics Sciences, University of São Paulo, Ribeirão Preto, Brazil, [3] Department of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of São Paulo, São Paulo, Brazil, [4] Department of Biochemistry and Immunology, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil, [5] Department of Medicine, University Hospital-Empresa Brasileira de Serviços Hospitalares (EBSERH), Federal University of Sergipe, Aracaju, Brazil, [6] Hospital Israelita Albert Einstein, São Paulo, Brazil, [7] Fiocruz-Bi-Institutional Translational Medicine Platform, Ribeirão Preto, Brazil

OPEN ACCESS

Edited by:
Xu-jie Zhou,
Peking University First Hospital, China

Reviewed by:
Bolaji N. Thomas,
Rochester Institute of Technology, United States
Hira Nakhasi,
Center for Biologics Evaluation and Research (FDA), United States
Nahid Ali,
Indian Institute of Chemical Biology (CSIR), India
Jose Carlos Solana,
Universidad Autónoma de Madrid, Spain

***Correspondence:**
Sandra Regina Maruyama
sandrarcm@ufscar.br

†These authors share senior authorship

Specialty section:
This article was submitted to Autoimmune and Autoinflammatory Disorders, a section of the journal Frontiers in Immunology

Received: *27 September 2021*
Accepted: *11 February 2022*
Published: *15 March 2022*

Citation:
Maruyama SR, Fuzo CA, Oliveira AER, Rogerio LA, Takamiya NT, Pessenda G, de Melo EV, da Silva AM, Jesus AR, Carregaro V, Nakaya HI, Almeida RP and da Silva JS (2022) Insight Into the Long Noncoding RNA and mRNA Coexpression Profile in the Human Blood Transcriptome Upon Leishmania infantum Infection. Front. Immunol. 13:784463. doi: 10.3389/fimmu.2022.784463

Visceral leishmaniasis (VL) is a vector-borne infectious disease that can be potentially fatal if left untreated. In Brazil, it is caused by *Leishmania infantum* parasites. Blood transcriptomics allows us to assess the molecular mechanisms involved in the immunopathological processes of several clinical conditions, namely, parasitic diseases. Here, we performed mRNA sequencing of peripheral blood from patients with visceral leishmaniasis during the active phase of the disease and six months after successful treatment, when the patients were considered clinically cured. To strengthen the study, the RNA-seq data analysis included two other non-diseased groups composed of healthy uninfected volunteers and asymptomatic individuals. We identified thousands of differentially expressed genes between VL patients and non-diseased groups. Overall, pathway analysis corroborated the importance of signaling involving interferons, chemokines, Toll-like receptors and the neutrophil response. Cellular deconvolution of gene expression profiles was able to discriminate cellular subtypes, highlighting the contribution of plasma cells and NK cells in the course of the disease. Beyond the biological processes involved in the immunopathology of VL revealed by the expression of protein coding genes (PCGs), we observed a significant participation of long noncoding RNAs (lncRNAs) in our blood transcriptome dataset. Genome-wide analysis of lncRNAs expression in VL has never been performed. lncRNAs have been considered key regulators of disease progression, mainly in cancers; however, their pattern regulation may also help to understand the complexity and heterogeneity of host immune responses elicited by *L. infantum* infections in humans. Among our findings, we identified lncRNAs such as IL21-AS1, MIR4435-2HG and LINC01501 and coexpressed lncRNA/mRNA pairs such as CA3-AS1/CA1, GASAL1/IFNG and LINC01127/IL1R1-IL1R2. Thus, for the first time, we present an integrated analysis of PCGs and lncRNAs by exploring the lncRNA–mRNA coexpression profile of VL to provide insights into the regulatory gene network involved in the development of this inflammatory and infectious disease.

Keywords: blood transcriptomics, human visceral leishmaniasis, *Leishmania infantum* (syn. *Leishmania chagasi*), mRNA sequencing (mRNA-seq), long noncoding RNA–mRNA coexpression

INTRODUCTION

Leishmania protozoans cause a group of diseases known as leishmaniases. The diseases are characterized by a wide range of clinical manifestations depending on the infecting *Leishmania* species, which are classified as cutaneous, mucocutaneous or visceral forms. This dixenous and dimorphic parasite is transmitted to humans through bites from infected sand flies (1). Infections by *Leishmania infantum* can lead to the most severe form of disease, visceral leishmaniasis (VL), which can be lethal if untreated or misdiagnosed (2). Human cases in Brazil account for approximately 95% of reported VL cases in the Americas, with a mortality rate of 7.2% (3). It is also classified as a zoonotic disease with dogs and wild animals as reservoirs. Antileishmanial treatment is administered only parenterally and triggers many toxicity effects, and currently, there is no vaccine against VL. The control of the vector and the surveillance of reservoirs of *L. infantum* have been the measures adopted by the Brazilian public health policies to control and prevent the disease (4). Patients susceptible to VL go through a minimum period of remission of six months, and recidivism has been more frequently observed in recent years. Primarily, children are more often affected, but the incidence in adults has significantly increased (5). However, as observed for other infectious diseases, most infected people do not become sick or even develop any symptoms, likely due to the diverse factors influencing the complexity of the host/parasite interface.

Most mechanistic knowledge about *Leishmania* infections arises from experimental animal models and eventually from human infections. *Leishmania* parasites are able to infect multiple cell types, of which mononuclear phagocytes are the main cells for intracellular replication. By establishing a long-term infection, the parasites are capable of escape from microbicidal and immune mechanisms. $CD4^+$ Th1 cells and IFN-γ production are crucial to control *Leishmania* infections, but other $CD4^+$ T cell subtypes and $CD8^+$ T cells have been shown to be important in the adaptive immune response (6, 7). Displaying a variety of *Leishmania* species, hosts and infection scenarios, the combinations of host/pathogen interactions reach a diversity of host responses to be investigated. In this context, studies aiming to uncover the molecular mechanisms underlying the different outcomes of *L. infantum* human infections are still scarce.

Blood transcriptomics represents an accessible and powerful approach to address the molecular immune mechanisms elicited by inflammatory conditions (8) and to foster the understanding of the heterogeneity of many human infectious diseases. In this regard, blood transcriptomes of human VL caused by *L. donovani* have been explored by others in India (9), with a focus on amphotericin B treatment. Another blood transcriptomics study using patients from Africa also focused on treatment efficacy assessment, but in VL patients coinfected with HIV (10). Of interest for VL occurrence in Brazil, a pioneering study performed by Gardinassi et al. with VL patients infected with *L. infantum* revealed molecular immunological signatures according to the outcome of infection and disease state, such as asymptomatic infection, active infection and during VL remission between two to five months after treatment with pentavalent antimonial (11). These studies found that the IFN-γ response circuit was enriched in active VL (as expected), pathways related to the activation of T lymphocytes *via* MHC class I, type I interferon signaling and B cells (11). Adriaensen et al. showed that IL-10 integrated a 4-gene pre-post transcriptional signature to discriminate treatment outcomes (10).

All these blood transcriptomics studies were dedicated to defining transcriptional signatures of protein-coding genes (PCGs). Another type of gene, as important as PCGs, is those classified as long noncoding RNAs (lncRNAs) owing to their key roles in several molecular processes, such as gene regulation (namely, posttranscriptional and posttranslational mechanisms), genome integrity, cellular structural functions and interference in signaling pathways (12). None of these previous transcriptome studies in VL have focused on lncRNAs. Long noncoding RNAs are broadly expressed in health and disease states, and their specific or altered expression profiles indicate their potential as biomarkers and targets for novel therapies. Most lncRNA functions and relevance came from studies with tumors, but their central role in hematopoiesis and immunity is quite prominent (13, 14).

This type of transcript is larger than 200 nucleotides, and like mRNA, it is spliced, capped at the 5' end and polyadenylated at the 3' end (15), i.e., it can be captured not only by total RNA sequencing but also by mRNA sequencing. From this perspective, we performed an integrated analysis of lncRNAs and mRNAs (PCGs) in the blood transcriptomes of human VL caused by *L. infantum* obtained by mRNA sequencing. To produce robust findings and overcome potential host genetics factors, we compared transcriptional data of the same patients in two defined states, during active VL and after six months of being treated, when they were considered clinically cured. In addition, we compared the gene expression profiles of active VL to two other non-diseased profiles, asymptomatic individuals and healthy uninfected volunteers. First, we provided an overview of this new blood transcriptome in VL depicting the most enriched biological pathways and the featured landscape of expressed lncRNAs. Then, we proceeded to the differential expression analysis to define gene subsets to be further focused on in lncRNA–mRNA coexpression analysis. We provided an expression profile of lncRNAs induced in VL during *L. infantum* infection associated with coexpressed protein coding genes, uncovering important insights into the transcriptional response of this parasitic infectious disease. Several lncRNAs were identified as key players in human *L. infantum* infections, and their potential as blood biomarkers for VL is discussed.

MATERIAL AND METHODS

Patients and Healthy Uninfected Subjects

Twenty-nine individuals were enrolled in the study and categorized into four groups: visceral leishmaniasis patients with active disease (PD0), cured VL patients (PD180; VL patients from the PD0 group 180 days after treatment), asymptomatic individuals (A) and healthy uninfected controls (C), as summarized in **Figure 1** and **Table 1**. The individuals enrolled in this study are from Sergipe state, located at the Northeast region of Brazil, that is not endemic for Malaria. All procedures were performed in accordance with the guidelines of the Brazilian Human Research Ethics Evaluation System (CEP/CONEP) and were approved by the Ethics Committee of the Federal University of Sergipe (CAAE: 04587312.2.0000.0058). All subjects or their legal guardians signed an informed consent form prior to the study.

Diseased patients (PD0 samples) were characterized by the presence of fever, weight loss, hepatosplenomegaly, and low leukocyte and platelet counts. VL diagnosis in the PD0 group was confirmed by direct observation of *Leishmania* in bone marrow aspirate or positive culture in Novy–MacNeal–Nicolle (NNN) medium plus a positive rK39 serological test (Kalazar Detect Rapid Test, InBios International Inc., Seattle, WA). All VL patients were negative for hepatitis B and C viruses and HIV, and also for bacterial infections or other parasites. The patients were treated with conventional drug therapies used for visceral leishmaniasis, according to the national guidelines of the Brazilian Ministry of Health: meglumine antimonate (Glucantime®) and/or

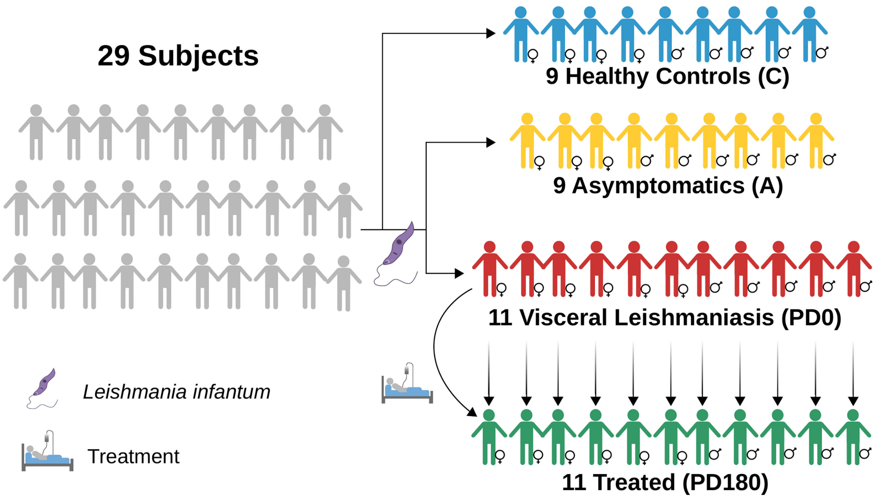


FIGURE 1 | Diagram depicting the groups used in this work to perform RNA-seq data analyses of blood transcriptomes from VL patients (PD0, in red) compared to nondiseased groups, treated (PD180 in green, cured patients), asymptomatic subjects IgG$^+$ for *Leishmania infantum* (A, in yellow) and healthy/control subjects (C, in blue). Image diagrammed in Inkscape (https://inkscape.org/).

TABLE 1 | General information of the groups analyzed in this study.

Group	N°	Women			Men		
		N°	Age[b,c]	Drug therapy[d]	N°	Age[b,c]	Drug therapy
PD0[a] (VL)	11	6	14.5 (01/51)	Glucantime + AmBisome (n = 2); AmBisome (n = 3)	5	24 (10/44)	Glucantime + AmBisome (n = 2); AmBisome (n = 2); Glucantime (n = 1)
PD180[a] (cured VL)	11	6	14.5 (01/51)		5	24 (10/44)	
Asymptomatic	9	3	12 (03/30)	–	6	21 (06/42)	–
Healthy Control	9	4	25.5 (24/27)	–	5	23 (11/29)	–

[a] *Same patients before (diseased) and after treatment (cured).*
[b] *Ages are expressed as the mean values (in years) and minimum and maximum values in parentheses (min/max).*
[c] *Age variance among groups was not statistically significant (Kruskal–Wallis test, p-value = 0.5157).*
[d] *Not available for one patient.*

liposomal amphotericin B (AmBisome®). In the follow-up appointment, after 180 days of VL treatment, the patients were considered clinically cured, comprising the PD180 group (totaling 22 paired samples, n = 11 in each time point). Healthy individuals who presented normal hematologic indices and neither clinical signs nor symptoms of VL but positive reactions to leishmanial antigens (Montenegro Skin test and rK39 serological test) were considered asymptomatic (n = 9), i.e., they were infected by *L. infantum* but without development of the disease. Parasite visualization tests were not performed in the PD180 group or asymptomatic individuals. Healthy individuals with negative tests for leishmanial antigens comprised the control group (n = 9).

Blood Sample Collection and RNA Isolation

Peripheral blood samples were collected using BD Vacutainer® tubes for hematologic tests and PAXgene Blood RNA tubes for RNA isolation. Total RNA was extracted from whole blood with the PAXgene Blood RNA Kit followed by globin mRNA depletion using the GLOBINclear™ Human Kit to enrich the samples for RNA from leukocytes. RNA samples were checked for purity by absorbance measurements (nm) of the 260/280 and 260/230 ratios using a NanoDrop™ 1000 Spectrophotometer and quantified using a Qubit™ 3.0 Fluorometer with a Qubit™ RNA HS Assay Kit. Assessment of RNA quality was obtained with RIN values >7.0 (RNA Integrity Number) with an Agilent 2100 Bioanalyzer using a Bioanalyzer RNA 6000 Nano assay.

mRNA Sequencing (mRNA-seq)

mRNA-seq data were generated in Illumina sequencing technology at the Genomics Center of the Laboratory of Animal Biotechnology, ESALQ, University of São Paulo, Piracicaba, Brazil, following the workflow recommended by the instructions of the manufacturer. Polyadenylated cDNA libraries were prepared with 300 µg of RNA depleted from globin mRNAs using the TruSeq® Stranded RNA Sample Preparation Kit. Paired-end sequencing was performed using a HiSeq SBS V4 kit (2 × 100 and 2 × 125 reads) in a HiSeq 2500 sequencer, yielding approximately 71 million reads for each mRNA-seq library.

RNA-seq Data Analysis

Raw fastq files were checked for quality control using FastQC (16). Illumina sequencing adaptors were trimmed, and low-quality reads (Phred score lower than 20, Q20) were filtered out using Trimmomatic (17). Read mapping was performed with STAR aligner (18) using the human genome reference assembly GRCh38.p38 (provided by The Genome Reference Consortium) annotated by the Ensembl database (19). Concordant uniquely mapped reads were used for downstream analyses. Quantification of reads to gene features used the –quantMode GeneCounts function from STAR. Read counts were used for differential expression analyses with the edgeR package in R (20), applying a quasi-likelihood F test (glmQLFTest function) with batch effect correction. A threshold false discovery rate lower than 0.05 (FDR <0.05) and a cutoff of 2-fold regulation (−1< log2-fold-change >+1) were used to fill the differentially expressed gene (DEG) list for each possible comparison between groups. TPM (transcripts per kilobase million) values were obtained by dividing read counts by the mean length of each gene in kilobases achieved by GTFtools to obtain the reads per kilobase (RPK) (21) and then dividing the RPK values by the sum of all RPK values in millions in a sample. Gene annotations were retrieved from Ensembl using the biomaRt R package (22).

Modular gene coexpression analyses were performed with the CEMiTool R package (23) using embedded functions for gene set enrichment analysis (GSEA) and overrepresentation analysis (ORA) with pathways from the Reactome database (24). The sample heterogeneity of gene expression profiles was assessed by the Molecular Degree of Perturbation (MDP) R package (25). Cell type composition based on blood RNA-seq data was predicted by CIBERSORT (26), a cellular deconvolution method. Long noncoding RNA (lncRNA) gene annotation was performed using the Ensembl BioMart (https://www.ensembl.org/biomart/martview/) (27) and the LNCipedia database (https://lncipedia.org/) (28). The functional genomics public repositories GEO/NCBI (29) and ArrayExpress/EMBL-EBI (30) were used to search other *Leishmania*-related transcriptomes, and three blood transcriptomics studies published elsewhere were selected for comparative analyses of detected differentially expressed long noncoding RNAs (DE lncRNAs): GSE77528 (11), GSE125993 (9) and PRJNA595895 (10).

Prioritized DEGs for lncRNA-mRNA coexpression analysis were obtained by overlapping the DEG lists using a Venn diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/). The proportions of lncRNAs and mRNAs in the DEG results were calculated by the chi-square test using the chisq.test function in R and plotted with the corrplot R package (31). Pearson's correlation implemented in R base functions was used to find coexpressed pairs of DE lncRNA-mRNA based on prioritized DEG lists encompassed by 147 lncRNAs and 1,263 protein coding genes (PCG, mRNA molecules) that were differentially expressed; only coexpressed pairs with resulting correlation coefficients of −0.8< *r* >0.8 were used for downstream analysis regarding lncRNAs. Network analysis of lncRNA–mRNA pairs was performed using igraph (32) and ggnetwork R packages (33). Hub genes were identified by betweenness and centrality measures. Subcellular localization of lncRNAs was retrieved from the LncSLdb database (34) and/or predicted using the LncLocator webtool (35). Interactions of lncRNAs with other molecules were searched or predicted with RNAInter (36). In the case of interaction with miRNAs, mRNAs targeted by miRNAs were searched using the miRDB database (37).

Statistical Analysis for the Demographic and Clinical Parameters

Statistical analysis of demographic (**Table 1**) and hematological (**Table 2**) data of the twenty-nine individuals was performed using GraphPad Prism v5.02 software. Each measured parameter was preliminarily assessed for normality using D'Agostino & Pearson and Shapiro–Wilk tests. Student T test was used to compare two groups, if the data follow normal distribution. Mann–Whitney U test was used to compare two groups, if the

TABLE 2 | Hematological data of the groups analyzed in this study.

Parameter	PD0—VL diseased (mean ± SD)	PD180—VL cured (mean ± SD)	Asymptomatic (mean ± SD)	Control (mean ± SD)	*p-value* for comparisons*
RBC (10^6/mm^3)	3.47 ± 0.35	5.11 ± 0.68	5.16 ± 0.537	4.61 ± 0.37	<0.001^b (1, 2 and 3)
Hemoglobin (g/dl)	7.98 ± 1.01	12.2 ± 3.6	13.95 ± 2.128	13.23 ± 0.25	<0.01^a (1, 2 and 3)
Hematocrit (%)	26.55 ± 5.58	40.6 ± 5.4	42.511 ± 4.9	39.85 ± 1.59	<0.01^b (1, 2 and 3)
Platelets (10^3/mm^3)	148.18 ± 81.63	239.9 ± 57.1	277.44 ± 144.53	253.75 ± 73.67	<0.05^a (1 and 2)
WBC (10^3/mm^3)	3,500.1 ± 1,948.8	7,114.5 ± 1,051.2	6,421.1 ± 1,423.7	6,260 ± 1161.2	<0.05^a (1, 2 and 3)
Neutrophils (10^3/mm^3)	1,329.5 ± 1,318.7	3,224.9 ± 753.3	3,286.7 ± 423.5	3,422.5 ± 258.505	<0.01^b (1, 2 and 3)
Eosinophils (10^3/mm^3)	51.00 ± 123.84	676.9 ± 531.2	332.8 ± 175.98	213.2 ± 144	<0.01^b (1 and 2)
Basophils (10^3/mm^3)	10.10 ± 21.76	98.8 ± 100.2	84.8 ± 43.4	87 ± 40.4	<0.001^b (1, 2 and 3)
Lymphocytes (10^3/mm^3)	1,644.70 ± 861.50	2,643.4 ± 956.2	2,273.9 ± 1,256.8	2,065 ± 626.5	<0.01^a (1)
Monocytes (10^3/mm^3)	360.20 ± 186.72	470.5 ± 186.6	445.2 ± 215.09	473.5 ± 141.7	NS^a

**Comparisons: 1 = PD180 vs PD0; 2 = Asymptomatic (A) vs PD0; 3 = Control (C) vs PD0; 4 = A vs PD180; 5 = A vs C; 6 = C vs PD180.*

[a] p-value calculated by t-Test (paired t-test for PD180 vs PD0).

[b] p-value calculated by Mann–Whitney test (Wilcoxon signed rank test for PD180 vs PD0).

NS, non-significant.

data followed non-Gaussian distribution. To compare the paired groups, PD0 and PD180, paired T-test and Wilcoxon signed rank test were used for data with normal and non-Gaussian distribution, respectively. Age variance in multiple groups was tested using Kruskall–Wallis followed by Dunn's post-test for non-Gaussian distribution. *P*-values lower than 0.05 were considered for statistical significance.

RESULTS AND DISCUSSION

Features of the Polyadenylated Transcriptome of Blood From Visceral Leishmaniasis Patients

We generated and analyzed bulk RNA-seq data of whole blood samples collected from 11 patients with active VL ("P–D0" abbreviations) and six months after the treatment, when these individuals were considered clinically cured ("P–D180" abbreviations). To gather insights into not only the immunopathophysiology of the disease but also the molecular mechanisms involved in *L. infantum* infection, we enrolled nine asymptomatic individuals (positive for anti-*L. infantum*, "A" abbreviations) and nine healthy subjects as control individuals ("C" abbreviations). Our analyses consisted of a total of 40 RNA-seq libraries depleted from globin and poly(A)+ selected mRNA, yielding an average of 4 Gb and 24 million paired-end mapped reads per library. The main characteristics of the analyzed groups and the clinical laboratory data for VL patients are displayed in **Tables 1** and **2**, respectively.

After filtering out genes with low expression (cpm <2; N <3) from the dataset, a total of 14,247 genes remained to be further analyzed. A multidimensional scaling (MDS) plot of the gene dataset was built to visualize the similarity across the 29 individuals represented by 40 samples. As shown in **Figure 2A**, there was clear segregation between active VL (PD0 group) and VL-free (PD180, A and C groups) individuals. Apart from P27 patient, all VL patients presented noticeably distinct gene expression patterns when considered clinically cured of the disease (PD180 group), which clustered together with other VL-free groups, asymptomatic (A) and healthy uninfected controls (C). Despite being within the same cluster, patients with active VL presented dissimilarities among them, reflecting the multifaceted nature of the disease. The heterogeneity of these blood transcriptomes was also assessed by molecular degree of perturbation (MDP) scores (25), in which PD0 samples presented the highest scores as expected, but it is also interesting to note that among VL-free groups, asymptomatic and cured patients presented distinct scores from healthy uninfected people (**Figure 2B**).

For an overview of the system-level functionality of all genes expressed in blood during the development of visceral leishmaniasis, we performed modular gene coexpression analysis using CEMiTool (23). Eleven different coexpressed modules were identified in the Gene Set Enrichment Analysis (GSEA), out of which 10 presented at least four significantly enriched pathways in the Over Representation Analysis (ORA). In the GSEA plot (**Figure 2C**, right panel), the activity of each module is displayed for all studied groups. Among the modules enriched across all groups, we depicted the modules M2, M5 and M6 with high Normalized Enrichment Scores (NES) that were only active in the VL group (PD0), which presented enriched pathways related mainly to "O2/CO2 exchange in erythrocytes", "Interferon Signaling" and "Cell Cycle Checkpoints", respectively. The crucial role of IFN- γ signaling in leishmaniasis (38) and the key roles of type I interferons in protozoan infections, including *Leishmania*, have been increasingly established in recent years (39). Modules M1 and M8 were related mainly to the pathways "neutrophil degranulation" and "Toll-like receptor cascades", respectively, and presented a low NES (activity) in active VL (**Figure 2C**, right panel). Neutrophils are massively recruited upon *Leishmania* infections, but the parasite efficiently evades neutrophil killing (40). They also influence the different forms of leishmaniasis, but they have been reported to play either protective or harmful roles during infection, depending on *Leishmania*-infecting species (41). The neutrophil response to infection outcomes seems to depend on their recruitment phase and tissue environment (42). Toll-like receptors (TLRs) are another innate immune branch acting in *Leishmania* infections, with multiple TLRs being activated simultaneously, and the interplay among them may

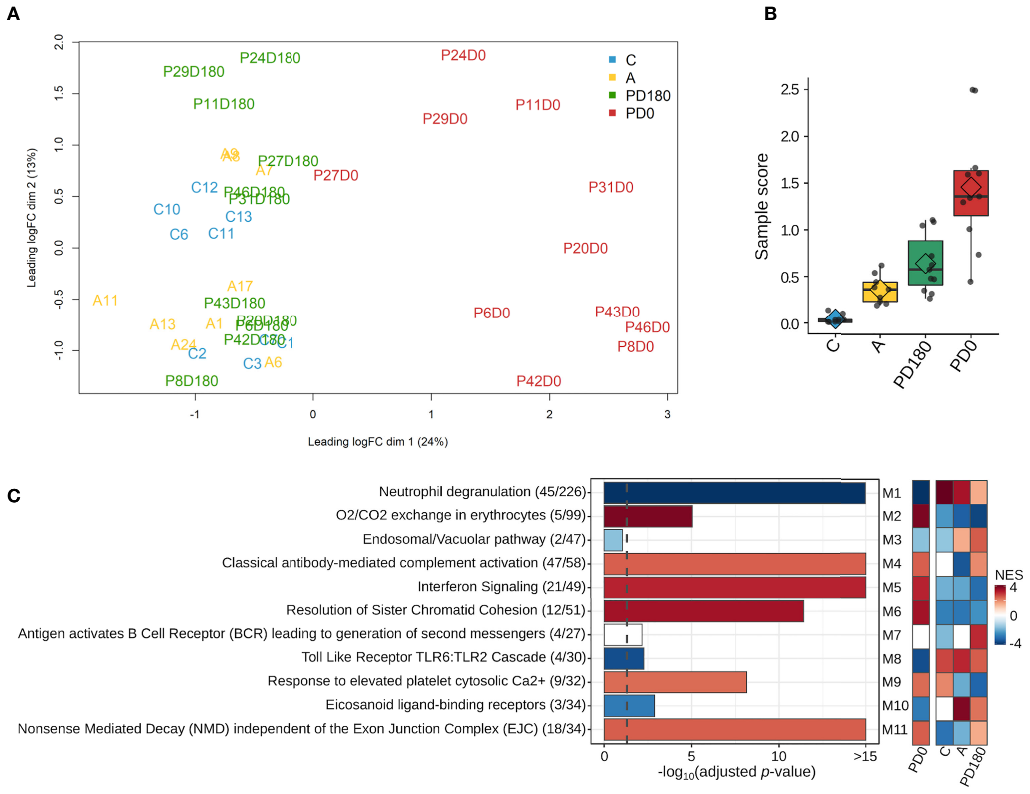


FIGURE 2 | Overview of the blood transcriptomes of this study. A total of 14,247 genes were expressed across the four groups, which encompassed 40 samples from 29 subjects. **(A)** Multidimensional scaling (MDS) plot and **(B)** molecular degree of perturbation (MDP) plot of all expressed genes in the four groups; **(C)** most significantly enriched pathways retrieved from the Reactome database for each module. The gene ratio for each pathway is displayed in parentheses. Colors in the Reactome enrichment graph refer to module activity in VL patients (PD0 group) as represented by the GSEA plot (heatmap in the right panel) displaying the module activity for each group. Color intensity is proportional to NES (normalized enrichment score). The graded scale side bar (NES) from red to blue indicates higher and lower activity, respectively, based on the ranked expression level.

influence the final outcome of infection (43). In an experimental model of *L. infantum* infection, TLR2 was shown to be important in promoting a protective immune response and effector mechanism of neutrophils (44). In addition, the TLR4 and IFN-I pathways play significant roles in preventing chronic inflammatory processes and immunopathology during *L. infantum* infection (45). All significantly enriched pathways elicited in VL are visualized in **Presentation 1** within each module.

Next, we proceeded to differential expression analysis to identify genes (DEGs) that were significantly regulated. The lists of DEGs (cutoff: FDR <0.05 and log2-fold-change ± 1) for pairwise comparisons can be found in **Supplementary Table 1**. As observed by others (9, 11), no DEG was found between asymptomatic and healthy control individuals considering the *post hoc* test at FDR<0.05. Considering that both groups basically include healthy and VL-free individuals, it is coherent do not finding difference statically significant between them. We focused on comparisons of the three non-diseased groups, A, C and PD180, against the active VL group, PD0. The comparisons presented an average of 1,680 DEGs each, with many hundreds of up- or downregulated genes (**Figures 3A–C**). All three comparisons presented some dozens of highly regulated genes (−3< log2-fold-change >3) with very statistically significant FDR values (FDR <0.0001), as can be observed at the superior corners of the volcano plots. Among many DEGs, we highlighted PRSS33 (serine protease 33), which was upregulated in the nondiseased groups (i.e., downregulated during active VL) at least 32-fold. PRSS33, also known as EOS, was primarily identified as expressed predominantly by macrophages, and also in peripheral leukocytes, and was detected in many organs, such as spleen, intestine, lung and brain (46). More recently, it was found that the production of PRSS33 by leukocytes is attributed specifically to eosinophils, which present constitutive expression at the mRNA level and cell surface expression at the protein level rather than being secreted (47). Interestingly, the gene expression pattern of PRSS33 along with IL10, SLFN14, and HRH4 was identified as a transcriptional signature to assess the treatment efficacy of visceral leishmaniasis in HIV patients (10). Here, the eosinophil count in VL patients was significantly decreased during *L. infantum* infection (**Table 2**); interestingly, the eosinophil gene signature was only detected in asymptomatic individuals (**Figure 4B**). The important role of eosinophils during

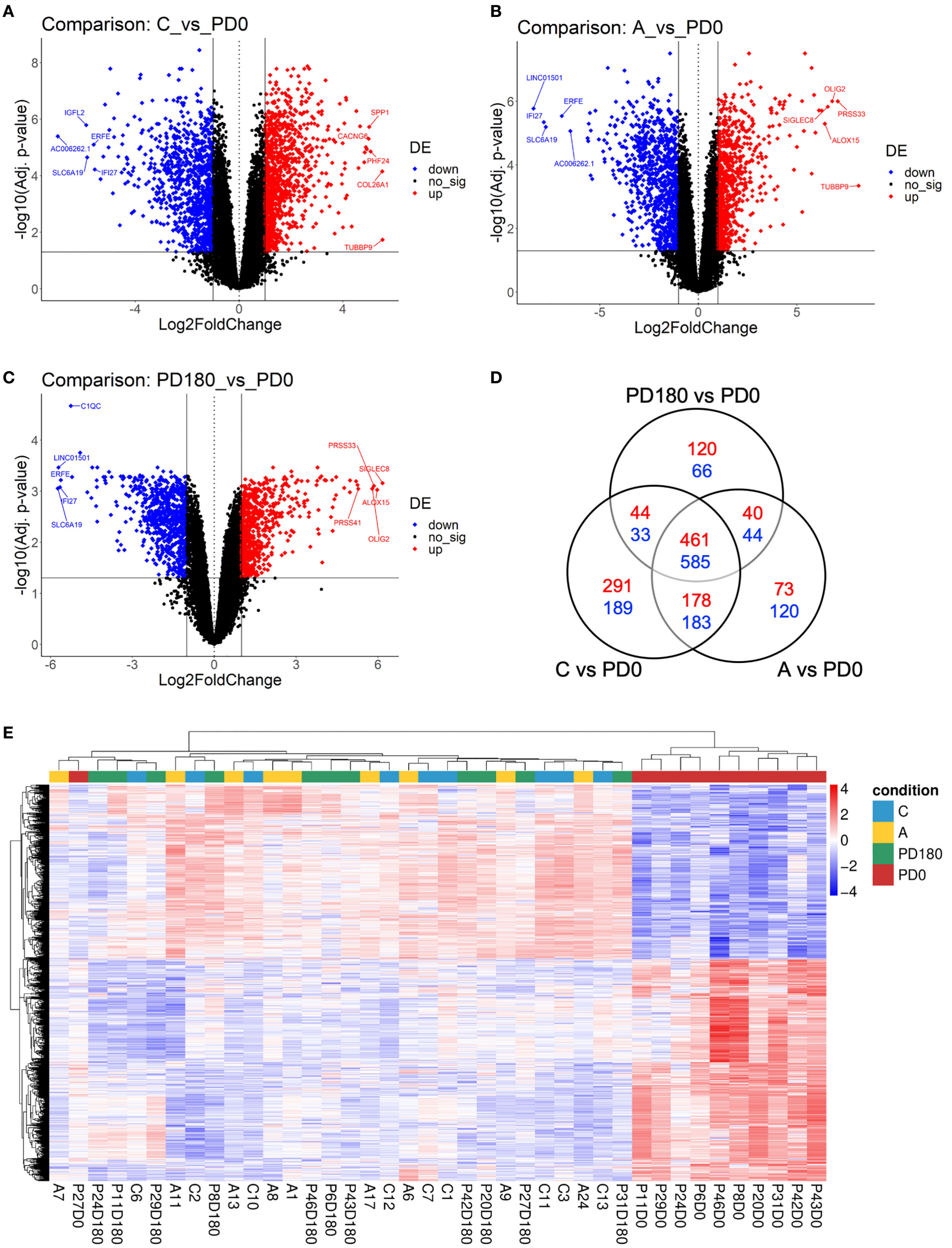


FIGURE 3 | Differential expression analysis comparing VL patients to nondiseased patients identified hundreds of differentially expressed genes (DEGs). Volcano plots highlighting genes significantly regulated (FDR <0.05, horizontal threshold line set on y-axis) in the comparisons C vs. PD0 **(A)**, A vs. PD0 **(B)** and PD0 vs. PD180 **(C)**. Vertical lines at -1 and +1 on the x-axis indicate the expression level criteria of fold decrease or increase, respectively, applied to DEGs that were further analyzed (all genes colored in blue or red); **(D)** Venn diagram displaying the number of exclusive DEGs for each comparison, as well as the number of shared DEGs among them. Dashed squares indicate the DEG lists used as subsets of prioritized genes in the mRNA-lncRNA coexpression profile analysis; **(E)** Heatmap of 1,045 DEGs shared among the three comparisons (suggested to be the gene signature of VL disease status), depicting the clustering of samples in two major groups: VL patients (PD0 labels, in red), the very consistent cluster on the right side and a heterogeneous cluster encompassing the nondiseased groups (**C, A** and PD180 labels, in blue, yellow and green, respectively) split into minor clusters. Z scores of cpm read counts were used, and a graded color scale from red to blue indicates whether the level of gene expression was above or below the mean (i.e., up- or downregulation).

Leishmania infection has become increasingly evident since many studies have demonstrated that eosinophils are able to control parasite load and interact with innate and adaptive immune responses, mainly by shaping macrophage responses (48–50). Another interesting DEG is interferon alpha inducible protein 27 (IFI27), which was highly downregulated in the non-diseased groups (i.e., upregulated during active VL) by an average of 64-fold. IFI27 (also known as ISG12) is a gene highly induced by type 1 interferon with pro-apoptotic effects (51) and antiviral activity (52, 53) and a potential biomarker identified through transcriptomics in some cancers, such as pancreatic adenocarcinoma (54).

A Venn diagram of these three comparisons (**Figure 3D**) displayed the number of shared or exclusive DEGs. Considering all intersections, when the redundancy was filtered out, we identified 2,427 unique DEGs. The central overlap presented 461 upregulated and 584 downregulated genes, which represents the "disease" gene set, because regardless of the non-disease group, these genes were significantly regulated in the disease group. An unsupervised clustering of this disease status gene signature (1,045 DEGs) showed the substantial grouping of PD0 patient samples (**Figure 3E**). In addition, we depicted the exclusive gene sets of PD180 vs. PD0 and A vs. PD0 comparisons, and the intersection between them, which accounted for 40 up- and 44 downregulated genes and may reveal genes related to a "molecular footprint" of the *L. infantum* infection because both cured patients and asymptomatic individuals had already been infected (unrelated to whether they became sick or not). The exclusive gene set of PD180 vs. PD0 may indicate genes associated with "molecular scarring" triggered by immunopathological mechanisms of the disease. Last, the exclusive gene set of A vs. PD0 comparison (73 up- and 120 downregulated genes) might uncover genes related to infection control and resistance mechanisms, which abrogate the development of visceral leishmaniasis. To assign these signatures regarding "molecular footprint", "molecular scarring" and "controlling of infection" is difficult due to the individual sample heterogeneity in non-diseased groups (as observed in **Figure 2B** and **Supplementary Figure 1** for individual samples) but is still a valid assumption since the groups PD180 (cured) and Asymptomatic (A) groups presented higher molecular perturbation scores than the healthy control group, as assessed by the MDP tool (**Supplementary Figure 1**).

Furthermore, to assess the composition of cellular subtypes of leukocytes in VL, we applied the CIBERSORT method for deconvolution analysis of blood transcriptomes (**Figure 4**). In addition to standard subtypes observed in blood count data (**Table 2**), the leukocyte gene signature matrix was able to discriminate natural killer (NK) cells, and subsets of T and B cells. Through data reuse, we also compared our cellular profile with the composition of another study in VL with the blood transcriptome obtained by microarray (11) (**Supplementary Figure 2**). The cell proportions in whole blood RNA-seq data of VL patients (PD0) showed variations mainly in neutrophils, macrophages, monocytes, NK cells, T lymphocytes and plasma cells (**Figure 4A**). When compared to the study of Gardinassi et al. (**Supplementary Figure 2**), in general, similarities in alteration tendency between both transcriptomes regarding cellular composition but also unmatched alterations for some cell subtypes were observed. In addition to the composition, the proportions of most cell types (irrespective of which group) were

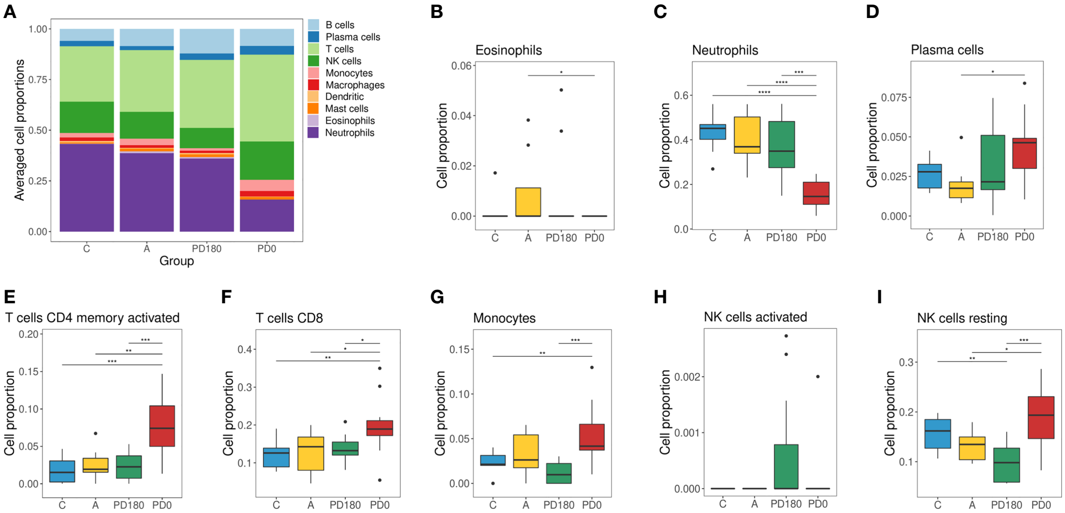


FIGURE 4 | Cellular deconvolution of blood transcriptomes in human visceral leishmaniasis using the CIBERSORT method: **(A)** Leukocyte proportions inferred from gene expression profiles of blood samples. Plots by cell type displaying the relative cell proportions for eosinophils **(B)**, neutrophils **(C)**, plasma cells **(D)**, activated memory CD4 cells **(E)**, CD8 T cells **(F)**, monocytes **(G)**, activated NK **(H)** cells and resting NK cells **(I)**. Groups were classified as active VL (PD0) and VL-free (PD180, A and C groups), in which the PD180 subjects were treated and considered clinically cured after 180 days of disease follow-up (PD0 and PD180 are paired groups, n = 11); A: asymptomatic (n = 9); C: healthy uninfected controls (n = 9). The differences between cell proportions were evaluated by Wilcoxon with Holm's correction. *P < 0.05, **P < 0.01, and ***P < 0.001.

different between the two datasets, such as neutrophils, plasma cells (both higher in this study) and monocytes (lower in this study). The profiles for neutrophils, plasma cells, activated memory CD4 T cells and CD8 T cells presented similar tendencies in both studies (**Figures 4C–F**), whereas the profile for monocytes presented the opposite behavior; here, it was increased in VL, but in Gardinassi et al., diseased patients presented a reduction in monocytes (**Figure 4G**). However, the variation in the number of monocytes before (PD0) and after treatment (PD180) was not significant, as shown in **Table 2**. This increase checked through CIBERSORT is related to the gene expression profile signed by monocytes. For NK cells, the profile presented complementary findings for activated (**Figure 4H**) and resting (**Figure 4I**) states between both studies, but both studies presented higher proportions of NK cells in active VL than in posttreatment, asymptomatic and control groups (**Figure 4A**). NK cells are able to recognize and are activated by *Leishmania* lipophosphoglycan (LPG), which is a dominant promastigote-specific surface glycoconjugate (55), *via* TLR-2 (56), but recently, it has been reported that human NK cells cannot be straightforwardly activated by *Leishmania* promastigotes and require monocyte-derived signals, such as transpresentation of IL-18, for their activation (57).

Long Noncoding RNA (lncRNA) Expression in Blood Upon *L. infantum* Infection

Several processed lncRNAs are capped and polyadenylated (15). Due to this feature, poly-A selection as a strategy of enrichment used in our RNA-seq was able to reveal the set of lncRNAs expressed in the blood of visceral leishmaniasis patients. Therefore, we also addressed the analysis of long noncoding RNAs, which is completely new in the VL transcriptomics field. From the total of 14,247 expressed genes, we identified 1,147 transcripts annotated as lncRNAs, according to gene biotype annotations in BioMart Ensembl. They were widely distributed across human chromosomes, from 1 to 22 and X, in which chromosomes 17 and 1 accounted for the largest numbers of lncRNAs, 95 (~8.3%) and 94 (~8.2%), respectively (**Figure 5A** and **Supplementary Table 2**). No lncRNA from the Y chromosome was detected in our RNA-seq dataset. The average gene length of the identified lncRNAs was 311,117 bp, with 9.3% being shorter than 1,000 bp and 41% being higher than 10,000 bp (**Figure 5B** and **Supplementary Table 2**). According to our analyses, 504 (~44%) lncRNAs were found to be expressed as unique transcripts (one transcript count in annotated genome version). Approximately 28% of lncRNAs had a transcript count from 2 to 5 (28%), and one expressed lncRNA (ENSG00000179818) presented 239 transcripts (**Figure 5C** and **Supplementary Table 2**). We identified the class of sequence ontology terms of 1,140 from 1,147 lncRNAs, namely, 484 (~42%) antisense, 482 (~42%) intergenic, 116 (~10%) sense intronic, 51 (~4.5%) bidirectional, and 7 (~0.6%) sense-overlapping lncRNAs (**Figure 5D** and **Supplementary Table 2**).

Based on the 2,427 unique DEGs found in the comparisons of nondiseased groups (A, C, and PD180) versus active VL, we searched for lncRNA gene biotype annotations in specific gene sets presented earlier (1,512 DEGs), focused on the central intersection (1,045 common DEGs) and the intersections between PD180 vs. PD0 and A vs. PD0 comparisons (88 DEGs), and their exclusive gene sets, which account for 186 and 193 DEGs, respectively (**Figure 3B**). A total of 147 lncRNAs (9.7%) out of 1,512 DEGs were found to be differentially expressed in this dataset (**Supplementary Table 3**), in which 89 (60.5%) and 58 lncRNAs were up- and downregulated, respectively, in the non-diseased groups compared to PD0 patients with active VL. Of note, the PD180 vs. PD0 exclusive gene set, with 186 DEGs, included 35 lncRNAs (19%), among which the expression of 30 genes significantly increased upon the clinical cure of VL, which might suggest that the suppression of these lncRNAs is related to the transcriptional regulation of protein-coding genes (PCGs) involved in the immunopathology of disease.

Notably, we highlighted MALAT1 (metastasis-associated lung adenocarcinoma transcript 1), which presented a statistically significant 2-fold increase in cured VL patients (PD180 group, **Supplementary Table 1**). MALAT1 is one of the most studied lncRNAs, exhibiting a variety of molecular regulatory functions in transcription and alternative splicing by binding in chromatin regions and binding in a plethora of protein, miRNA and mRNA molecules (58). MALAT1 has been extensively studied not only in oncology but also in many inflammatory diseases, where it plays a controversial role due to its action as either an oncogene or a tumor suppressor gene depending on the type of cancer (59). Interestingly, this controversial role of MALAT1 was also observed in parasitic protozoan infections, where deficiency of this lncRNA was important to enhance immunity and for clearance of *L. donovani* in a VL mouse model; however, in an experimental malaria model, *Malat1*$^{-/-}$ mice presented more severe disease (60). In this latter cited work, the authors suggest that MALAT1 is a nonredundant regulator of immunity by promoting the expression of the Maf/IL-10 axis in effector CD4^{+} T cells. This immune regulator function of MALAT1 was also found in tolerized mice with cardiac allografts by inducing tolerogenic dendritic cells and regulatory T cells through the miRNA-155/DC-SIGN/IL10 axis (61).

Although lncRNAs have portrayed less than a tenth of the polyadenylated transcriptome, the proportion of lncRNAs in the DEG gene set and their regulation pattern (up- or downregulation) were significantly associated and enriched in the differential expression data as calculated by the chi-square test (p-value = 0.0004, **Figure 5E**). The clustering of expression data of the 147 differentially expressed lncRNAs highlights the distinct expression pattern of selected lncRNAs in the active VL group (PD0, **Figure 6**). Generally, lncRNAs are expressed at low levels (62). Notably, some lncRNAs presented relatively high levels of expression, such as LINC01871 and AC012368.1 (**Figure 6**, green heatmap), as indicated by their TPM values. Other lncRNAs presented marked fold change regulation, such as IL21-AS1 and AC111000.4, which presented an average 16-fold increase ($\log_2$FC = 4) and an average 32-fold decrease ($\log_2$FC = −5), respectively, in the PD0 active VL group compared to the non-diseased groups (**Supplementary Table 3**). The most upregulated DE lncRNA in VL patients was the completely unknown lncRNA LINC01501 (long intergenic nonprotein coding RNA 1501), which presented an average fold

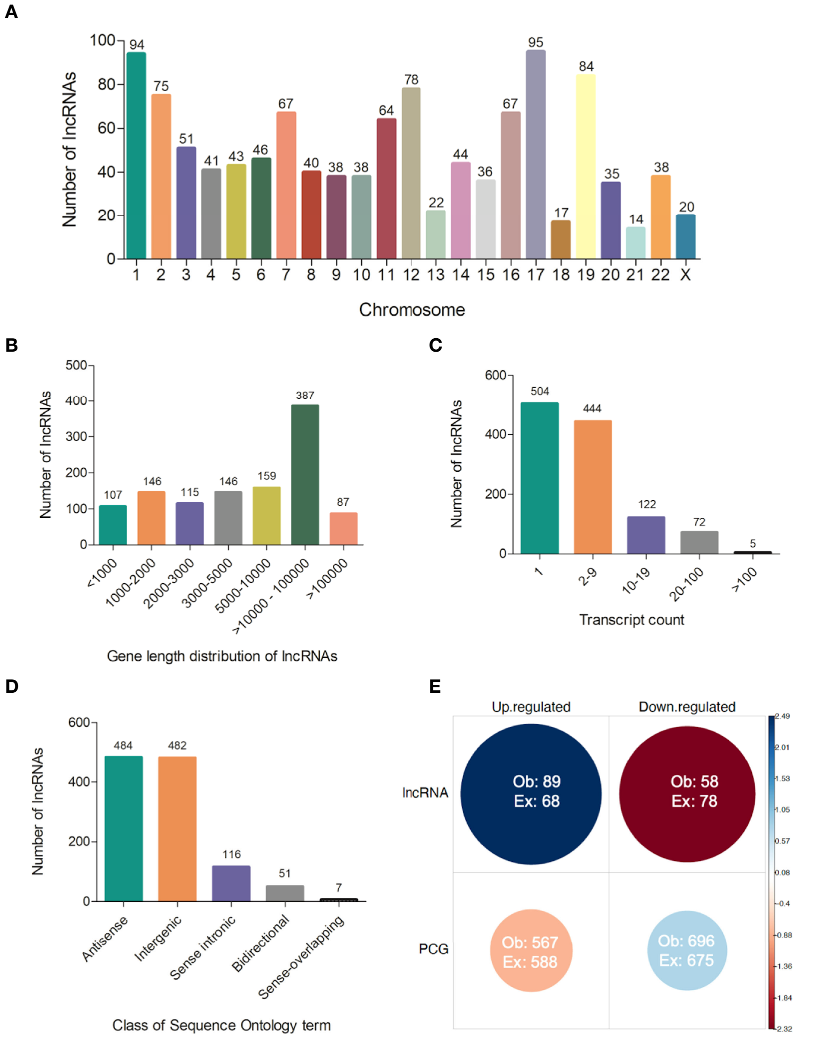


FIGURE 5 | Features of lncRNAs detected in polyadenylated RNA-seq of human visceral leishmaniasis. **(A)** Chromosomal distribution of lncRNAs across the 22 autosomes and the X chromosome; **(B)** Gene length distribution of lncRNAs; **(C)** Number of transcripts presented by each lncRNA; **(D)** Classification of lncRNAs according to sequence ontology terms; **(E)** Pearson's chi-square test using the number of lncRNAs and protein-coding genes (PCGs) and regulation patterns (up- or downregulated transcripts) in DEG datasets (comparisons of non-diseased groups, A, C and PD180 to PD0 active VL). The size and color intensity of circles are proportional to the contribution of the cell to the significance of the chi-square test. The standardized residuals are scaled at the sidebar, where positive and negative values indicate positive and negative associations, respectively. Ob, observed value; Ex, expected value, χ^2 = 12.343, df = 1, *p-value* = 0.0004.

increase of 64× ($\log_2$FC = 6) in the PD0 group (**Supplementary Table 4**).

lncRNA–mRNA Coexpression Analysis

Subsequently, we integrated the profiles of protein-coding genes (i.e., mRNAs) and long noncoding RNAs by performing lncRNA–mRNA coexpression analysis. Pearson correlations for the lncRNA–mRNA coexpression profile encompassing the 147 lncRNAs (89 up- and 58 downregulated) and the 1,263 (567 up- and 696 downregulated) mRNAs differentially expressed (as numbered in **Figure 5D**) identified 4,901 positive and 1,223 negative highly correlated pairs of lncRNA–mRNA associations (**Supplementary Table 5**). Networks were built using these coexpression correlations (**Figure 7**) to identify hubs of lncRNAs (e.g., KDM7A-DT, USP30-AS1 and LINC01501) and mRNAs (e.g., NPRL3, LAG3 and E2F2). The top 20 hubs within these networks for lncRNAs and mRNAs were identified by flagging them with their respective gene symbols. The top pairs of coexpressed lncRNA–mRNA positively (e.g., AC111000.4-CCR3; AC111000.4-IL5RA) and negatively correlated profiles (e.g., USP30-AS1-CCN3; LINC01501-IL1RAP) were identified by ranking Pearson's correlation results (**Supplementary Table 5**). The main pathways enriched by this highly correlated lncRNA–mRNA expression profile corroborated the results found when the whole expression profiling was analyzed (**Figures 2C, D** and **Presentation 1**), with common Reactome enrichment results (**Supplementary Table 5**).

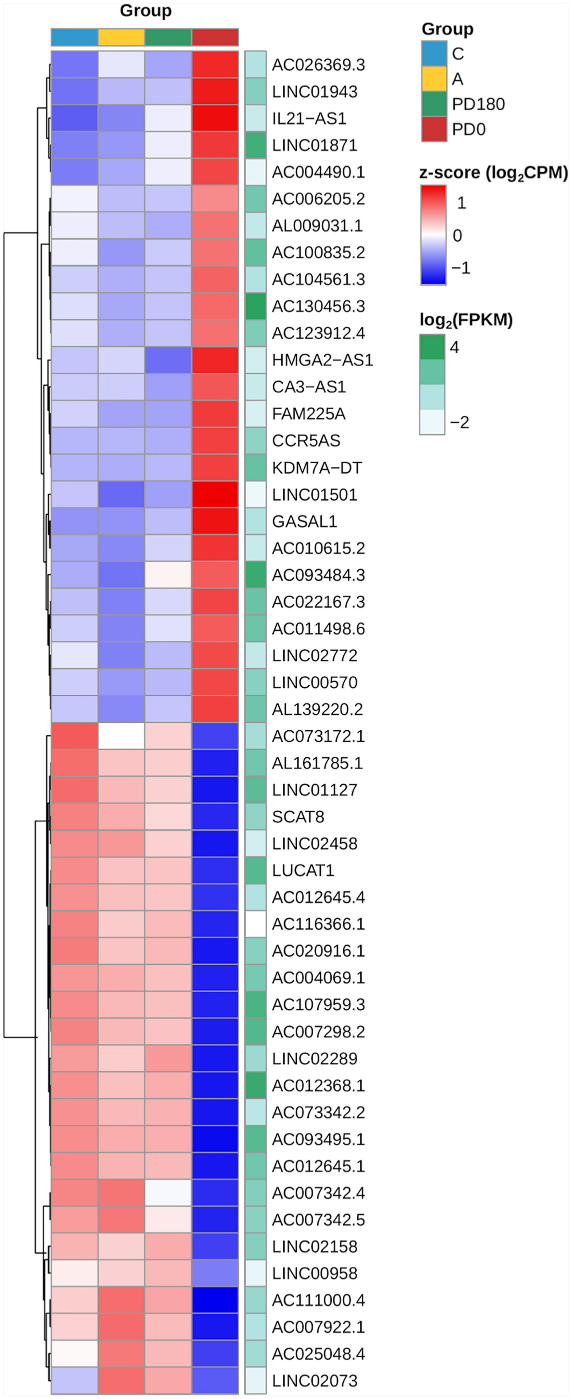


FIGURE 6 | Heatmap of the expression profiles of the top 20 lncRNAs expressed in VL patients (PD0). The top selection retrieved those genes within the dataset of 147 primarily selected DE lncRNAs (**Supplementary Table 2**) based on ranking of statistical significance (FDR values) followed by fold regulation ($\log_2$FC values). Gene expression regulation by group is represented by the z score of the cpm read count, and the transition color scale from red to blue indicates up- and downregulation, respectively. Gene expression level is represented by mean FPKM values at the side (green heatmap column), where color intensity toward dark green indicates increasing levels (highly expressed lncRNAs).

Regulatory networks of lncRNAs act in *cis*- and *trans*-regulation. Human genome annotation has revealed that the vicinity of PCGs is surrounded by lncRNAs, and as we verified by sequence ontology of expressed lncRNAs, most of them were classified as antisense or intergenic (**Figure 5D**). *Cis*-acting lncRNAs play gene regulation functions from their own transcription sites, operating in PCGs at proximal distances within the same chromosome (63). Many intergenic (lincRNA) long noncoding RNAs are placed in topologically associated domains (TADs), an approximately 1 Mb genomic segment featuring chromatin interactions; *cis*-acting lincRNAs have been associated with modeling the chromosomal architecture (64). To infer potential *cis*-regulation, the highly correlated lncRNA–mRNA pairs were tracked into their genomic coordinates. lncRNAs within a 300-kb window size from the transcription start site (TSS) of the correlated PCG were retrieved, resulting in 22 potential *cis*-acting lncRNAs all positively associated with their respective mRNAs (**Figure 8**). The majority are located downstream of the TSS of the PCG pair, and the five sites upstream of the PCG pair are lincRNAs. Out of six potential *cis*-acting lncRNAs upregulated in VL patients, CA3-AS1 (CA3 antisense RNA 1) presented higher fold regulation than the non-diseased group ($\log_2$FC = 3.035, approximately 9-fold increase). CA3-AS1 has been identified as a key lncRNA in gastrointestinal cancers (65–67). Even more differentially expressed, its correlated PCG pair, CA1 (carbonic anhydrase 1), presented a 24-fold increase during active *L. infantum* infection. GO (Gene Ontology) biological process classification for CA1 revealed that this protein participates in the interleukin-12-mediated signaling pathway (GO:0035722) and is involved in gene and protein expression by JAK-STAT signaling after IL-12 stimulation according to the Reactome database (R-has-8950505).

As shown by genomic localization mapping, most coexpressed lncRNA–mRNA pairs were located in different chromosomes or distally in the same chromosome, making the majority of lncRNA networks hypothetically play a role in transcription as a *trans*-acting regulation. In fact, lncRNAs that may act near their transcription sites may also undertake regulatory functions far from the TSS or even outside the nucleus (12, 63). Mechanisms for the transregulation of transcription depend on interactions of lncRNAs with proteins, DNA and other RNA molecules. Additionally, depending on their subcellular localization, lncRNAs may interfere with posttranscriptional and intracellular signaling (15). Based on this framework, we proceeded to mine data for lncRNA subcellular localization and lncRNA interactions using the LncSLdb database (34) and the LncLocator webtool (35) for subcellular localization and the RNAInter database (36) for an interaction overview of hub lncRNAs selected from coexpression network analyses (**Figure 7** and **Supplementary Table 5**), and inferred *cis*-acting lncRNAs (**Figure 8**). The analysis of this subset comprises 51 lncRNAs from the 147 DE lncRNA list (**Supplementary Table 4**) detailed in our work. A piece of this interactome is summarized in **Table 3**, and interactions related to the 51 lncRNAs are available as additional material (**Supplementary Table 6**).

Most of the interactors found in RNAInter is a database belong to microRNA (miRNA) and transcription factor

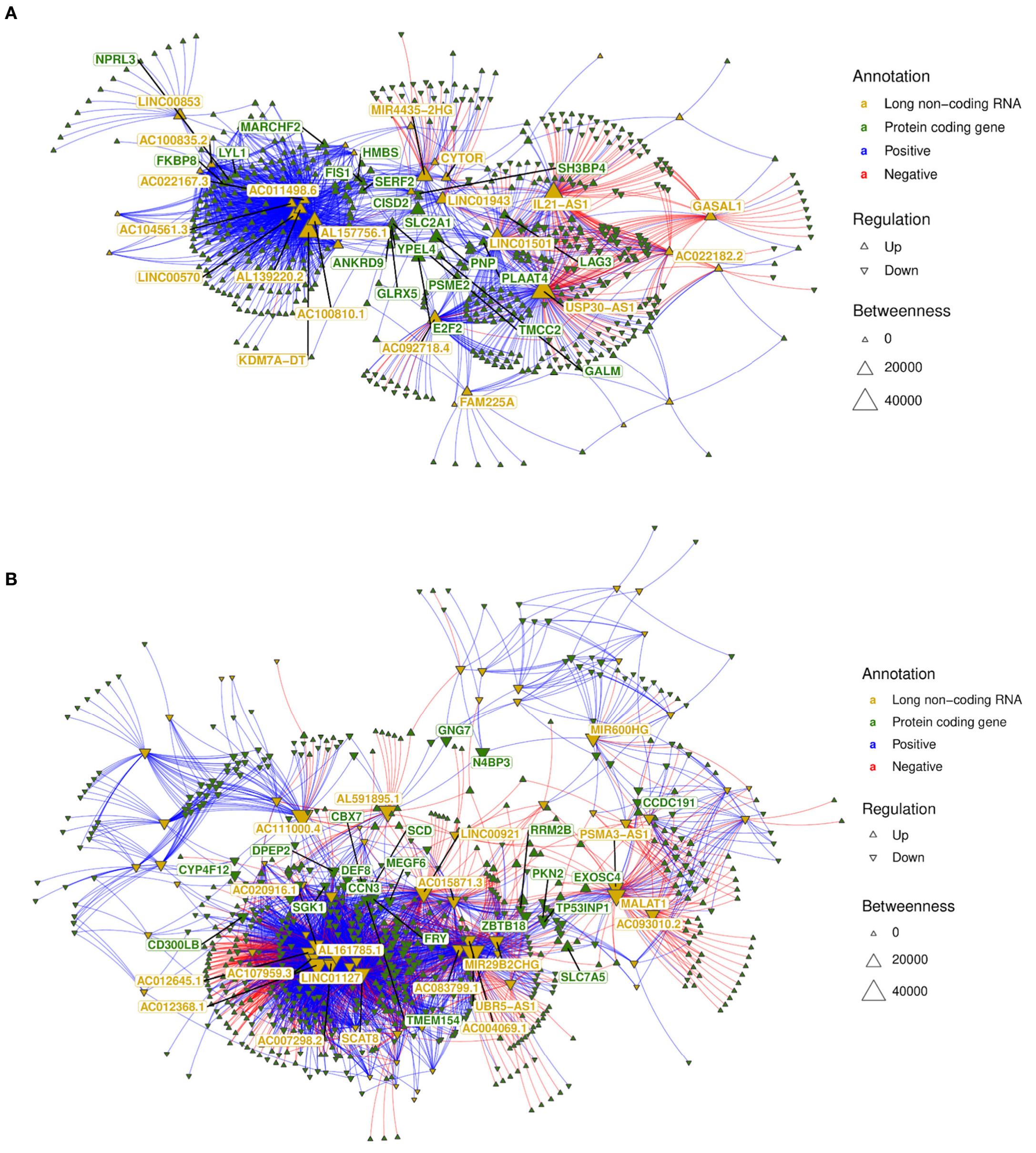


FIGURE 7 | lncRNA–mRNA coexpression network analysis. **(A)** A network was built with all highly correlated lncRNA–mRNA pairs (1,870 pairs; **Supplementary Table 6**) obtained for 58 lncRNAs downregulated in non-diseased groups (i.e., upregulated in VL patients—PD0); **(B)** A network was built with all highly correlated lncRNA–mRNA pairs (4,256 pairs; **Supplementary Table 6**) obtained for 89 lncRNAs upregulated in nondiseased groups (i.e., downregulated in VL patients—PD0). The top 20 hubs for lncRNAs and mRNAs are flagged with gene symbols in yellow and green, respectively.

categories. Of note, 8 out of 51 lncRNAs did not present any annotation in noncoding RNA databases (**Supplementary Table 6**) but reached high scores for cytoplasm/cytosol subcellular localization prediction. Three lncRNAs of these 8 (AL157756.1, AC100810.1, and AC007922.1) presented exclusively at least 40 highly positively coexpressed PCGs (i.e., same pattern regulation) and three other lncRNAs presented mixed correlations of PCG pairs, including AC111000.4, the most downregulated lncRNA in VL patients. For those interactions with miRNA, we searched for miRNA targets using miRDB (37) to provide the PCGs coexpressed with the respective lncRNAs, which in turn were retrieved from coexpression network analysis (last column of **Table 3**). lncRNAs that present miRNA binding sites can interact with miRNAs by acting as competing endogenous RNAs (ceRNAs) or natural miRNA sponges, building another complex layer of the transcriptional regulatory network (68). As shown in **Table 3**, CA3-AS1 interacts with hsa-miR-93-5p, which in turn targets the NEDD4L (E3 ubiquitin-protein ligase NEDD4-like) protein and the transcriptional repressor MXI1 (Max-interacting

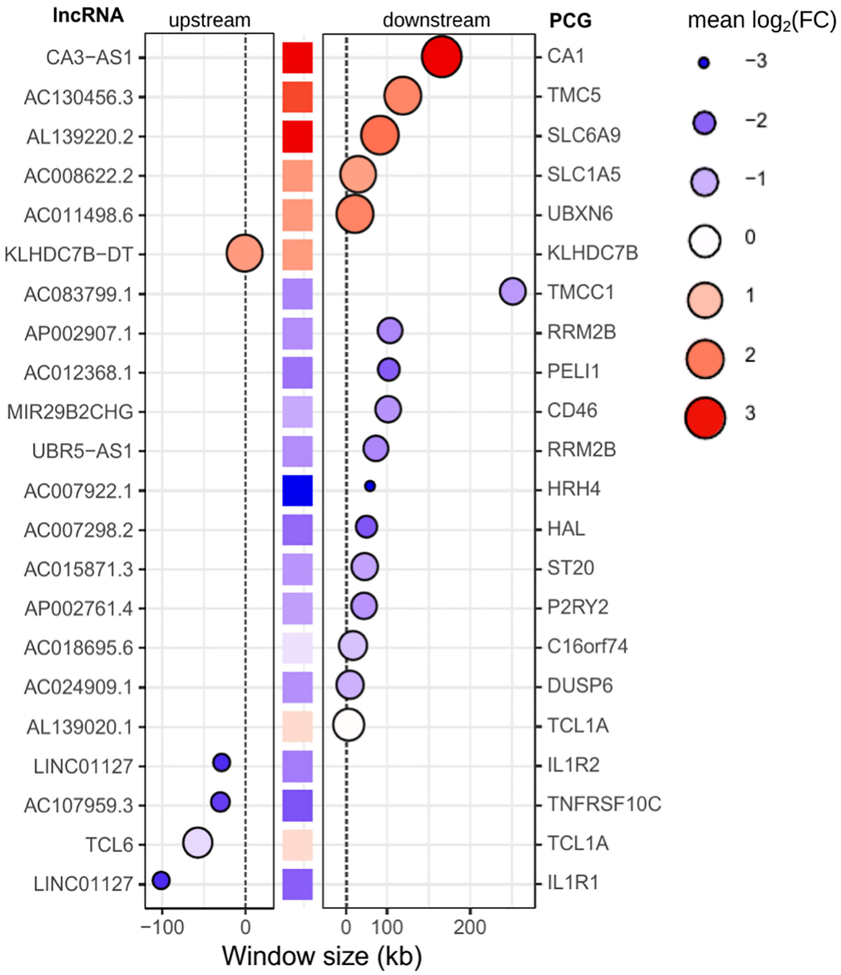


FIGURE 8 | Potential *cis-acting* lncRNAs inferred from highly correlated lncRNA-mRNA pairs in the genomic vicinity. A window size of 300 kb was set to consider *cis*-regulation. Gene IDs on the left refer to lncRNAs, whereas gene IDs on the right refer to PCGs (mRNAs). The central column of squares indicates the position of PCGs related to lncRNAs, and the graded color indicates the pattern regulation in PD0 (logFC). Bubbles indicate the position of lncRNAs upstream (left panel) or downstream (right panel) of the PCG. The size and color intensity of bubbles indicate the pattern regulation of lncRNAs in PD0. The blue scale indicates downregulation, whereas the red scale indicates upregulation in VL patients (PD0).

protein 1); CA3-AS1, NEDD4L and MXI1 were upregulated during *L. infantum* infection. The ceRNA regulator function of CA3-AS1 was described elsewhere (66), where it was found to be an anti-oncogene in gastric cancer by sponging hsa-miR-93-5p. Interactions with miRNAs were found for the other 6 lncRNAs, of which 5 interacted with multiple miRNAs. The upregulated hub GASAL1 (growth arrest-associated lncRNA 1) presented the highest number of miRNA interactions and was the only lncRNA correlated with IFNG expression. GASAL1 has been shown to be involved in the inhibition of tumor growth in lung cancer and may improve chronic heart failure by downregulating the TGF-β signaling pathway (69, 70).

The upregulated lncRNAs, AL139220.2 (novel transcript) and KDM7A-DT (KDM7A divergent transcript), were predicted to interact with the same five transcriptional regulators that were coexpressed with LYL1 (Protein lyl-1), KLF1 (Kruppel-like factor 1), PBX1 (Pre-B-cell leukemia transcription factor 1), TAL1 (T-cell acute lymphocytic leukemia protein 1) and MXI1 (Max-interacting protein 1), with the latter being a repressor of target genes. PBX1 is involved in natural killer cell differentiation (GO:0001779), whereas LYL1 plays a role in B cell differentiation (GO:0030183). KLF1 and TAL1 are transcription regulators of hemopoietic differentiation. KDM7A-DT, also known as JHDM1D-AS1, has been suggested to play a protective role during ROS-induced apoptosis in periodontal ligament stem cells (71). Experimental validation data for AL139220.2 were not found.

LYL1 and PBX1 were also interactors of the downregulated PSMA3-AS1 (PSMA3 antisense RNA 1), a lncRNA that presented predicted interactions with 4 miRNAs, including hsa-miR-105-5p, which targets TLR10 (downregulated in VL patients). PSMA3-AS1 was a DE lncRNA only in the comparison of PD180 vs. PD0 (**Supplementary Table 3**). Recently, PSMA3-AS1 has been validated as a ceRNA involved in the malignant phenotypes of esophageal cancer by modulating the miR-101/

TABLE 3 | Potential lncRNA subcellular localization and interactors based on highly correlated lncRNA–mRNA pairs differentially expressed during *L. infantum* infection.

Potential *cis*-acting

lncRNA	PCG (mRNA) pair in chromosome vicinity	Number of PCGs coexpressed (total; positive; negative)[b]	Subcellular localization[c]	Number and type of interaction[d,e]	Coexpressed PCG pairs[f]
↑ AL139220.2[a]	↑ SLC6A9	128; 128; 0	Cytosol[c2]	4 miRNA: hsa-miR-107, hsa-miR-103a-3p, hsa-let-7c-5p, hsa-let-7b-5p 5 TF	miRNA targets: CARM1, ANK1, IFIT1B, TRIM10, PSMF1, FRMD4A, TAL1, TSPAN5, SERF2, XK, CDC34, RBM38, BCL2L1, PBX1, IGF2BP2 LYL1, KLF1, PBX1, MXI1, TAL1
↑ CA3-AS1	↑ CA1	22; 22; 0	–	1 miRNA: hsa-miR-93-5p 1 TF	miRNA targets: NEDD4 L, MXI1 MXI1
↑ AC130456.3	↑ TMC5	38; 38; 0	–	2 TF	LYL1, TAL1
↓ UBR5-AS1	↓ RRM2B	166; 100; 66	–	5 TF	↓ MXD1, ↓ BACH1, ↑ E2F2, ↑ FOXM1, ↑ EZH2
↓ AC012368.1	↓ PELI1	269; 172; 97	–	5 TF and 1 histone	↓ MXD1, ↓ FOS, ↑ E2F8, ↑ E2F7, ↑ EZH2, ↑ CENPA (histone)
↓ LINC01127	↓ IL1R1 ↓ IL1R2	214; 171; 43	–	4 TF 1 DNA	↓ MXD1, ↓ FOS, ↑ E2F7, ↑ EZH2 ↑ NUF2
↓ AC107959.3	↓ TNFRSF10C	214; 160; 54	–	5 TF and 1 histone	↓ MXD1, ↓ FOS, ↑ E2F8, ↑ E2F7, ↑ EZH2, ↑ CENPA (histone)

Potential *trans*-acting

lncRNA	Top coexpressed PCG pairs[g]	Number of PCGs coexpressed (total; positive; negative)[b]	Subcellular localization[c]	Number and type of interaction[d,e]	Coexpressed PCG pairs[f]
↑ FAM225A	↑ C2 ↑ FBXO6 ↑ PSME2	16; 16; 0	Nucleus/Cytoplasm[c1] Cytosol[c2]	1 miRNA: hsa-miR-1-3p	miRNA target: UBE2L6
↑ KDM7A-DT	↑ UBB ↑ STMP1 ↑ RBM38	154; 152; 2	–	5 TF	↑ LYL1, ↑ KLF1, ↑ PBX1, ↑ MXI1, ↑ TAL1
↑ IL21-AS1	↑ FABP5 ↑ CDKN2A ↑ CTLA4	117; 47; 70	Cytoplasm[c2]	2 TF	↓ TLE3, ↑ EZH2
↑ AC092718.4	↑ CDCA3 ↑ NCAPH ↓ IRS2	87; 78; 9	–	4 TF and 1 histone	↑ E2F2, ↑ FOXM1, ↑ MYBL2, ↑ EZH2, CENPA (histone)
↑ MIR4435-2HG	↑ H2AJ, ↑ PTMS, ↓ HDAC9	49; 33; 16	–	6 miRNAs: hsa-miR-6754-5p, hsa-miR-128-3p, hsa-miR-1185-5p, hsa-miR-105-5p, hsa-miR-103b, hsa-miR-1-3p, 1 TF	miRNA targets: ↓ ZFP28, ↓ PDE3B, ↓ ATP2B1, ↓ HDAC9, ↓ EPHA4, ↓ MOB3B, ↓ ZNF677, ↓ TCTN1, ↓ MAP3K1, ↑ POMP, ↑ E2F2, ↑ EMC3 ↑ BATF
↑ GASAL1	↓ EEPD1, ↓ ADGRE3, ↑ IFNG	38; 5; 33	Cytosol[c2]	7 miRNAs: hsa-miR-93-5p, hsa-miR-519d-3p, hsa-miR-20b-5p, hsa-miR-20a-5p, hsa-miR-17-5p, hsa-miR-106b-5p, hsa-miR-106a-5p	miRNA targets: ↓ TMCC3, ↓ SORL1 e ↓ OLIG1
↓ AC004069.1	↑ DNAJA4, ↓ ST3GAL6	65; 53; 12	Cytosol[c1], Ribosome [c1], Nucleus[c1] Cytoplasm[c2]	2 miRNAs: hsa-miR-10b-5p, hsa-miR-10a-5p 1 TF	miRNA targets: ↓ BAZ2B, ↑ NEDD4 L, ↑ ARHGEF12 ↓ MXD1
↓ PSMA3-AS1	↓ ZNF439 ↑ PSMF1	44; 19; 25	–	4 miRNAs: hsa-miR-106b-5p, hsa-miR-106a-5p, hsa-miR-101-3p, hsa-miR-105-5p 2 TF	miRNA targets: ↓ ZBTB18, ↓ ZFP28, ↓ PKN2, ↓ TP53INP1, ↓ ZFYVE16, ↓ VCPKMT, ↓ TMEM65, ↓ ZNF677, ↓ TLR10, ↑ TGFB1I1, ↑ PBX1 ↑ LYL1, ↑ PBX1

[a] *lncRNA hub in network analysis (**Figure 7**).*

[b] *Numbers were extracted from the expression correlation results available in **Supplementary Table 4**.*

[c] *Highlighted results from data mining in the LncSLdb Database (c1) or prediction by the LncLocator tool (c2). For in silico prediction, only scores higher than 0.7 (ranging from 0 to 1) were considered.*

[d] *Highlighted results from data mining in RNAInter Database. Interactions of lncRNAs with microRNAs (miRNAs), transcription factors (TFs), RNA binding proteins (RBPs), DNA and histone modifications; when lncRNAs exhibited interactions with miRNAs, targeted PCGs were listed.*

[e] *All lncRNAs displayed here present at least 35 histone modification results.*

[f] *PCGs were extracted from the expression correlation results available in **Supplementary Table 4**.*

[g] *PCGs with the highest correlation coefficients according to the results available in **Supplementary Table 4**.*

Up or down arrows indicate the expression regulation pattern in VL patients (PD0 group).

EZH2 axis (72). EZH2 (enhancer of zeste 2 polycomb repressive complex 2 subunit) is a histone methyltransferase and is another transcriptional repressor found as an interactor of our lncRNA network. In addition, PSMA3-AS1 has been discovered to sponge miR-409-3p and is considered an oncogenic lncRNA involved in the aggressive phenotype of non–small cell lung carcinoma (73).

Other coexpressed transcription factors commonly found as interactors of lncRNAs were the downregulated transcriptional repressor MXD1 (Max dimerization protein 1) and the E2F family of activators and repressors, whose coexpressed genes (E2F2, E2F7 and E2F8) were all upregulated in VL patients. The activity of E2F members is critical for transcriptional machinery throughout the cell cycle and cytokinesis (74). E2F2 is one of the transcription activators signed as a gene signature of the Th1 immune response (75). Finally, we highlight the upregulated lncRNA MIR4435-2HG, which was predicted to interact with 6 miRNAs, and BATF (basic leucine zipper ATF-like transcription factor), a transcription regulator that controls the differentiation of lymphocytes, specifically on switch isotypes in B cells and Th17 cells, follicular T-helper (TfH) cells, and CD8$^+$ dendritic cells by interacting with members of the interferon-regulatory factor (IRF) family (76). Moreover, BATF plays an essential role during hematopoiesis and the homeostasis of effector functions of innate lymphoid cells (ILCs) (77), which are innate counterparts of T cells and are predominantly situated at the mucosal barriers (78). To the best of our knowledge, this study is the first to infer a lncRNA–TF interaction between BATF and MIR4435-2HG within a transcriptional network regulation of a human infectious disease. Furthermore, MIR4435-2HG bound to EZH2 and promoted hepatocellular carcinoma progression *via* EZH2-mediated epigenetic silencing of p21 and E-cadherin expression (79). MIR4435-2HG lncRNA has been extensively studied in recent years, with 41 related articles in PubMed (https://www.ncbi.nlm.nih.gov/gene/541471), in which 34 of them have shown experimental validation of its role in tumor progression and as a prognostic biomarker in different types of cancer. It is also known as AGD2, LINC00978, MIR4435-1HG, MORRBID and lncRNA-AWPPH and acts as a ceRNA by sponging many other miRNAs (not listed in **Table 3**), such as miR-296-5p, which was identified to be part of the Akt2/SNAI1 signaling pathway involved in the development of oral squamous cell carcinoma triggered by *Fusobacterium nucleatum* infection (80).

Among the hub lncRNAs featured in our work, some of them have been shown to be stable and detectable in blood plasma samples and circulating exosomes elsewhere, such as MALAT1, MIR4435-2HG, and PSMA3-AS1 (81–83), which may promptly favor their potential as blood biomarkers. For other lncRNAs, such as IL21-AS1, AC111000.4, CA3-AS1, GASAL1, and LINC01501, no literature associated with plasma or circulating exosomes was found. However, since this study was not designed for diagnosis purpose, future studies should be performed to evaluate these lncRNAs as new avenue to be explored in VL. The significance of lncRNAs as master regulators of many biological processes in health and diseases is well established, and their application as biomarkers for disease progression has been rapidly increased in cancer biology but is still incipient in parasitic diseases. Furthermore, lncRNAs have been found in body fluids, freely or inside exosomes (84). Detecting eligible lncRNAs in plasma and/or serum is a promising noninvasive and affordable method for prognosis, and many studies with tumors have shown its value in either complementary diagnosis or aggressiveness prediction (84, 85).

CONCLUDING REMARKS

For the first time, an integrated analysis of lncRNAs and protein-coding genes (mRNAs) was performed in human visceral leishmaniasis using blood transcriptomics. Blood transcriptomes obtained by mRNA-seq allowed us to surpass the typical analyses comprising gene pathways and protein networks, adding an extra and important layer to this big picture captured by transcriptomics, the long noncoding RNA profile. From a comprehensive analysis, we highlighted lncRNAs such as MALAT1, CA3-AS1, GASAL1, PSMA3-AS1, MIR4435-2HG, IL21-AS1, AC111000.4, and LINC01501, with these last three being the most regulated lncRNAs compared active VL to the VL-free groups. Moreover, by comparing VL patients before and after a six-month follow-up, this study suggests there is a potential for use lncRNAs in plasma and/or serum as marker for monitoring disease remission. Focusing on a set of differentially expressed genes, the lncRNA–mRNA coexpression profile presented here was able to provide valuable and insightful data to help unravel the complexity of host/parasite interactions in human visceral leishmaniasis caused by *L. infantum* infection. We believe that our study will be useful to guide future studies for searching lncRNAs as biomarkers, new targets for drugs or drug repurposing and new therapies to control this neglected disease.

DATA AVAILABILITY STATEMENT

Metadata and fastq files from the RNA sequencing experiment were deposited in the functional 582 genomics repository, ArrayExpress from EMBL-EBI, under accession number E-MTAB-11047 583 (available at: http://www.ebi.ac.uk/arrayexpress/E-MTAB-11047/).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Brazilian Human Research Ethics Evaluation System (CEP/CONEP; CAAE: 04587312.2.0000.0058). Written informed consent to participate in this study was provided by own parcipants if in adult age or by the legal guardian/next of kin of the participants if under 18 years of age.

AUTHOR CONTRIBUTIONS

Conceived the study: RPA and JSS. Supervised the study: SRM, RPA, HN, and JSS. Designed the experiments: SRM, RPA, and

JSS. Generated clinical samples and management of clinical data: RPA, ARJ, EVM, and AMS. Performed the experiments: SRM and GP. Analyzed data: SRM, CAF, AERO, LAR, NTT, VC, and HN. Performed statistical data analysis: SRM, CAF, and AERO. Contributed reagents/material/analysis tools: SRM, ARJ, RPA, and JSS. Sketched figures and tables: SRM, LAR, NTT, CAF, and AERO. Wrote the manuscript: SRM. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was supported by grants from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP agreements 2016/20258-0 (Young Investigator Award to SM), 2018/14933-2 (HN), and 2013/08216-2 (Center for Research in Inflammatory Diseases); the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES) grant no. 23038.005304/2011-01 and Finance code 001; the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) grant no. 552721/2011-5 (RPA); fellowship/grant no. 307741/2017-6 to AJ. SM and AO received fellowships from the FAPESP (2017/16328-6 and 2019/22579-7, respectively). LR received a scholarship from the FAPESP (2018/26799-9 and 2020/14011-8), and NT received scholarships from the FAPESP (2021/12464-8) and the CNPq (133661/2020-2).

ACKNOWLEDGMENTS

We thank Iran Malavazi, Anderson Ferreira Cunha and Felipe Roberti Teixeira for their generous and continuing support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.784463/full#supplementary-material

Presentation 1 | CEMiTool html report with results for coexpression modules of RNA-seq analysis of blood in human visceral leishmaniasis caused by *L. infantum* infection.

Supplementary Figure 1 | Heterogeneity of samples analyzed using the **M**olecular **D**egree of **P**erturbation (MDP) tool. **(A)** Molecular perturbation scores plotted for each sample based on 14,247 genes expressed across the four groups. DEGs shared among all comparisons of VL patients to nondiseased groups (as numbers in **(Figure 3)**, central overlap); **(B–D)** Boxplots representing MDP scores by group quantified based on three subsets of DEGs, where the first was composed of 84 DEGs shared between A vs. PD0 and PD180 vs. PD0 comparisons (b), the second was composed of 186 DEGs exclusive for PD180 vs. PD0 comparison (c) and the third was composed of 193 DEGs exclusive for A vs. PD0 comparison.

Supplementary Figure 2 | Cellular deconvolution of blood transcriptomes in human visceral leishmaniasis using the CIBERSORT method. **(A)** Leukocyte proportions inferred from gene expression profiles of blood samples from Gardinassi et al., 2016 (microarray). **(B)** Plots by cell type displaying the relative cell proportions for neutrophils, eosinophils, plasma cells, memory CD4 cell plasma cells, CD8 T cells, monocytes, activated NK cells and resting NK cells. In the study of Gardinassi et al., groups were classified as CTRL (healthy uninfected controls, n= 15), DTH (asymptomatic patients, n= 14), TRT (patients at 2-5 months after treatment, considered under remission of the disease; n= 8) and VL (diseased patients, n= 8).

Supplementary Table 1 | Excel spreadsheets with results for differential expression analysis for pairwise comparisons between nondiseased groups PD180 (cured VL patients, 180 days after treatment), A (asymptomatic subjects) and C (healthy controls) against PD0 (active VL patients). Only differentially expressed genes (DEGs) at FDR <0.05 with -1< log2-fold-change > +1 (twofold decrease or increase) were included for further analysis.

Supplementary Table 2 | Annotation of 1,147 lncRNAs expressed in RNA-seq in this work. The main features were retrieved from the GRCh38.p13 human genome version using the BioMart Ensembl LNCipedia database.

Supplementary Table 3 | Annotation of 147 selected differentially expressed (DE) lncRNAs in RNA-seq in this work. The main features were retrieved from the GRCh38.p13 human genome version using the BioMart Ensembl, LNCipedia and Expression Atlas databases.

Supplementary Table 4 | Top50 lncRNAs selected from 147 DE lncRNAs (**Supplementary Table 3** spreadsheet) ranked by statistical significance (FDR values) followed by fold regulation ($\log_2$FC values) in VL patients compared to nondiseased groups.

Supplementary Table 5 | Highly correlated pairs of lncRNA-mRNA retrieved by Pearson's correlation using log(cpm) read count for 147 DE lncRNAs and 1,263 DE mRNAs from all samples. The cutoff for the correlation coefficient was -0.8 < *r* > 0.8; all results presented a *p value* < $6e^{-10}$. Hubs from networks and Reactome pathways are also tabulated herein.

Supplementary Table 6 | Catalog of lncRNA subcellular location and interaction partners searched for lncRNAs featuring hubs in network analysis **(Figure 7)** and predicted as *cis*-acting lncRNAs **(Figure 8)**. Data mining was performed using the LncSLdb and RNA Inter databases and the LncLocator prediction tool.

REFERENCES

1. WHO. *What Is Leishmaniasis?*. WHO. Available at: http://www.who.int/leishmaniasis/disease/en/ (Accessed August 24, 2018).
2. Hajj R, Hajj H, Khalifeh I. Fatal Visceral Leishmaniasis Caused by Leishmania Infantum, Lebanon. *Emerg Infect Dis* (2018) 24:906–7. doi: 10.3201/eid2405.180019
3. Alvar J, Vélez ID, Bern C, Herrero M, Desjeux P, Cano J, et al. Boer M Den, the WHO Leishmaniasis Control Team. Leishmaniasis Worldwide and Global Estimates of Its Incidence. *PloS One* (2012) 7:e35671. doi: 10.1371/journal.pone.0035671
4. Serafim TD, Iniguez E, Oliveira F. Leishmania Infantum. *Trends Parasitol* (2020) 36:80–1. doi: 10.1016/j.pt.2019.10.006
5. Lima ID, Lima ALM, Mendes-Aguiar C de O, Coutinho JFV, Wilson ME, Pearson RD, et al. Changing Demographics of Visceral Leishmaniasis in Northeast Brazil: Lessons for the Future. *PloS Negl Trop Dis* (2018) 12: e0006164. doi: 10.1371/journal.pntd.0006164
6. Kaye P, Scott P. Leishmaniasis: Complexity at the Host–Pathogen Interface. *Nat Rev Micro* (2011) 9:604–15. doi: 10.1038/nrmicro2608
7. Conceição-Silva F, Morgado FN. Leishmania Spp-Host Interaction: There Is Always an Onset, But Is There an End? *Front Cell Infect Microbiol* (2019) 9:330. doi: 10.3389/fcimb.2019.00330
8. Chaussabel D, Pascual V, Banchereau J. Assessing the Human Immune System Through Blood Transcriptomics. *BMC Biol* (2010) 8:84. doi: 10.1186/1741-7007-8-84

9. Fakiola M, Singh OP, Syn G, Singh T, Singh B, Chakravarty J, et al. Transcriptional Blood Signatures for Active and Amphotericin B Treated Visceral Leishmaniasis in India. *PloS Negl Trop Dis* (2019) 13:e0007673. doi: 10.1371/journal.pntd.0007673
10. Adriaensen W, Cuypers B, Cordero CF, Mengasha B, Blesson S, Cnops L, et al. Griensven J Van. Host Transcriptomic Signature as Alternative Test-of-Cure in Visceral Leishmaniasis Patients Co-Infected With HIV. *EBioMedicine* (2020) 55:102748. doi: 10.1016/j.ebiom.2020.102748
11. Gardinassi LG, Garcia GR, Costa CHN, Silva VC, Santos IKF de M. Blood Transcriptional Profiling Reveals Immunological Signatures of Distinct States of Infection of Humans With Leishmania Infantum. *PloS Negl Trop Dis* (2016) 10:e0005123. doi: 10.1371/journal.pntd.0005123
12. Statello L, Guo C-J, Chen L-L, Huarte M. Gene Regulation by Long Non-Coding RNAs and Its Biological Functions. *Nat Rev Mol Cell Biol* (2021) 22:96–118. doi: 10.1038/s41580-020-00315-9
13. Satpathy AT, Chang HY. Long Noncoding RNA in Hematopoiesis and Immunity. *Immunity* (2015) 42:792–804. doi: 10.1016/j.immuni.2015.05.004
14. Lüscher-Dias T, Conceição IM, Schuch V, Maracaja-Coutinho V, Amaral PP, Nakaya HI. Long Non-Coding RNAs Associated With Infection and Vaccine-Induced Immunity. *Essays Biochem* (2021) 65(4):657–69. doi: 10.1042/EBC20200072
15. Pinkney HR, Wright BM, Diermeier SD. The lncRNA Toolkit: Databases and In Silico Tools for lncRNA Analysis. *Non-Coding RNA* (2020) 6:49. doi: 10.3390/ncrna6040049
16. Babraham Institute software SA. *FastQC*. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
17. Bolger AM, Lohse M, Usadel B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* (2014) 30:2114–20. doi: 10.1093/bioinformatics/btu170
18. Dobin A, Gingeras TR. Mapping RNA-Seq Reads With STAR. *Curr Protoc Bioinf* (2015) 51:11.14.1-19. doi: 10.1002/0471250953.bi1114s51
19. Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, et al. Ensembl 2021. *Nucleic Acids Res* (2021) 49:D884–91. doi: 10.1093/nar/gkaa942
20. Robinson MD, McCarthy DJ, Smyth GK. Edger: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* (2010) 26:139–40. doi: 10.1093/bioinformatics/btp616
21. Li H-D. GTFtools: A Python Package for Analyzing Various Modes of Gene Models. *bioRxiv* (2018). doi: 10.1101/263517
22. Durinck S, Spellman PT, Birney E, Huber W. Mapping Identifiers for the Integration of Genomic Datasets With the R/Bioconductor Package biomaRt. *Nat Protoc* (2009) 4:1184–91. doi: 10.1038/nprot.2009.97
23. Russo PST, Ferreira GR, Cardozo LE, Bürger MC, Arias-Carrasco R, Maruyama SR, et al. CEMiTool: A Bioconductor Package for Performing Comprehensive Modular Co-Expression Analyses. *BMC Bioinf* (2018) 19:56. doi: 10.1186/s12859-018-2053-1
24. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res* (2020) 48:D498–503. doi: 10.1093/nar/gkz1031
25. Gonçalves ANA, Lever M, Russo PST, Gomes-Correia B, Urbanski AH, Pollara G, et al. Assessing the Impact of Sample Heterogeneity on Transcriptome Analysis of Human Diseases Using MDP Webtool. *Front Genet* (2019) 10:971. doi: 10.3389/fgene.2019.00971
26. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust Enumeration of Cell Subsets From Tissue Expression Profiles. *Nat Meth* (2015) 12:453–7. doi: 10.1038/nmeth.3337
27. Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, et al. Ensembl BioMarts: A Hub for Data Retrieval Across Taxonomic Space. *Database* (2011) 2011:bar030. doi: 10.1093/database/bar030
28. Volders P-J, Anckaert J, Verheggen K, Nuytens J, Martens L, Mestdagh P, et al. LNCipedia 5: Towards a Reference Set of Human Long Non-Coding RNAs. *Nucleic Acids Res* (2019) 47:D135–9. doi: 10.1093/nar/gky1031
29. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: Archive for Functional Genomics Data Sets—Update. *Nucleic Acids Res* (2013) 41:D991–5. doi: 10.1093/nar/gks1193
30. Athar A, Füllgrabe A, George N, Iqbal H, Huerta L, Ali A, et al. ArrayExpress Update – From Bulk to Single-Cell Expression Data. *Nucleic Acids Res* (2019) 47:D711–5. doi: 10.1093/nar/gky964
31. Wei T, Simko V. *R Package "Corrplot": Visualization of a Correlation Matrix. (Version 0.90), Https://Github.Com/Taiyun/Corrplot* (2021). Available at: https://github.com/taiyun/corrplot.
32. Csardi G, Nepusz T. *The Igraph Software Package for Complex Network Research* (2006). Available at: https://igraph.org.
33. Briatte F, Bojanowski M, Canouil M, Charlop-Powers Z, Fisher J, Johnson K, et al. *Ggnetwork: Geometries to Plot Networks With "Ggplot2"* (2021). Available at: https://cran.r-project.org/web/packages/ggnetwork/index.html.
34. Wen X, Gao L, Guo X, Li X, Huang X, Wang Y, et al. Lncsldb: A Resource for Long Non-Coding RNA Subcellular Localization. *Database* (2018) 2018: bay085. doi: 10.1093/database/bay085
35. Cao Z, Pan X, Yang Y, Huang Y, Shen H-B. The Lnclocator: A Subcellular Localization Predictor for Long Non-Coding RNAs Based on a Stacked Ensemble Classifier. *Bioinformatics* (2018) 34:2185–94. doi: 10.1093/bioinformatics/bty085
36. Lin Y, Liu T, Cui T, Wang Z, Zhang Y, Tan P, et al. RNAInter in 2020: RNA Interactome Repository With Increased Coverage and Annotation. *Nucleic Acids Res* (2020) 48:D189–97. doi: 10.1093/nar/gkz804
37. Chen Y, Wang X. miRDB: An Online Database for Prediction of Functional microRNA Targets. *Nucleic Acids Res* (2020) 48:D127–31. doi: 10.1093/nar/gkz757
38. Mirzaei A, Maleki M, Masoumi E, Maspi N. A Historical Review of the Role of Cytokines Involved in Leishmaniasis. *Cytokine* (2021) 145:155297. doi: 10.1016/j.cyto.2020.155297
39. Silva-Barrios S, Stäger S. Protozoan Parasites and Type I IFNs. *Front Immunol* (2017) 8:14. doi: 10.3389/fimmu.2017.00014
40. Regli IB, Passelli K, Hurrell BP, Tacchini-Cottier F. Survival Mechanisms Used by Some Leishmania Species to Escape Neutrophil Killing. *Front Immunol* (2017) 8:1558. doi: 10.3389/fimmu.2017.01558
41. Hurrell BP, Regli IB, Tacchini-Cottier F. Different Leishmania Species Drive Distinct Neutrophil Functions. *Trends Parasitol* (2016) 32:392–401. doi: 10.1016/j.pt.2016.02.003
42. Ribeiro-Gomes F, Sacks D. The Influence of Early Neutrophil-Leishmania Interactions on the Host Immune Response to Infection. *Front Cell Infect Microbiol* (2012) 2:59. doi: 10.3389/fcimb.2012.00059
43. Chauhan P, Shukla D, Chattopadhyay D, Saha B. Redundant and Regulatory Roles for Toll-Like Receptors in Leishmania Infection. *Clin Exp Immunol* (2017) 190:167–86. doi: 10.1111/cei.13014
44. Sacramento LA, Costa D LJ, Lima DFMH, Sampaio PA, Almeida RP, Cunha FQ, et al. Toll-Like Receptor 2 Is Required for Inflammatory Process Development During Leishmania Infantum Infection. *Front Microbiol* (2017) 8:262. doi: 10.3389/fmicb.2017.00262
45. Sacramento LA, Benevides L, Maruyama SR, Tavares L, Fukutani KF, Francozo M, et al. TLR4 Abrogates the Th1 Immune Response Through IRF1 and IFN-β to Prevent Immunopathology During L. Infantum Infection. *PloS Pathog* (2020) 16:e1008435. doi: 10.1371/journal.ppat.1008435
46. Chen C, Darrow AL, Qi J, D'andrea MR, Andrade-Gordon P. A Novel Serine Protease Predominately Expressed in Macrophages. *Biochem J* (2003) 374:97–107. doi: 10.1042/bj20030242
47. Toyama S, Naoko O, Matsuda A, Saito H, Nakae S, Karasuyama H, et al. A Novel Protease, PRSS33 (Serine Protease 33), Is Specifically and Constitutively Expressed in Eosinophils. *J Allergy Clin Immunol* (2017) 139:AB163. doi: 10.1016/j.jaci.2016.12.535
48. Rodríguez NE, Wilson ME. Eosinophils and Mast Cells in Leishmaniasis. *Immunol Res* (2014) 59:129–41. doi: 10.1007/s12026-014-8536-x
49. Lee SH, Chaves MM, Kamenyeva O, Gazzinelli-Guimaraes PH, Kang B, Pessenda G, et al. M2-Like, Dermal Macrophages Are Maintained *via* IL-4/CCL24–mediated Cooperative Interaction With Eosinophils in Cutaneous Leishmaniasis. *Sci Immunol* (2020) 5:eaaz4415. doi: 10.1126/sciimmunol.aaz4415
50. da Silva Marques P, da Fonseca-Martins AM, Carneiro MPD, Amorim NRT, de Pão CRR, Canetti C, et al. Eosinophils Increase Macrophage Ability to Control Intracellular Leishmania Amazonensis Infection *via* PGD2 Paracrine Activity *In Vitro*. *Cell Immunol* (2021) 363:104316. doi: 10.1016/j.cellimm.2021.104316
51. Gytz H, Hansen MF, Skovbjerg S, Kristensen ACM, Hørlyck S, Jensen MB, et al. Apoptotic Properties of the Type 1 Interferon Induced Family of Human Mitochondrial Membrane ISG12 Proteins. *Biol Cell* (2017) 109:94–112. doi: 10.1111/boc.201600034

52. Ullah H, Sajid M, Yan K, Feng J, He M, Shereen MA, et al. Antiviral Activity of Interferon Alpha-Inducible Protein 27 Against Hepatitis B Virus Gene Expression and Replication. *Front Microbiol* (2021) 12:656353. doi: 10.3389/fmicb.2021.656353
53. Brochado-Kith Ó, Martínez I, Berenguer J, González-García J, Salgüero S, Sepúlveda-Crespo D, et al. HCV Cure With Direct-Acting Antivirals Improves Liver and Immunological Markers in HIV/HCV-Coinfected Patients. *Front Immunol* (2021) 12:723196. doi: 10.3389/fimmu.2021.723196
54. Khatri I, Bhasin MK. A Transcriptomics-Based Meta-Analysis Combined With Machine Learning Identifies a Secretory Biomarker Panel for Diagnosis of Pancreatic Adenocarcinoma. *Front Genet* (2020) 11:572284. doi: 10.3389/fgene.2020.572284
55. Forestier C-L, Gao Q, Boons G-J. Leishmania Lipophosphoglycan: How to Establish Structure-Activity Relationships for This Highly Complex and Multifunctional Glycoconjugate? *Front Cell Infect Microbiol* (2015) 4:193. doi: 10.3389/fcimb.2014.00193
56. Becker I, Salaiza N, Aguirre M, Delgado J, Carrillo-Carrasco N, Kobeh LG, et al. Leishmania Lipophosphoglycan (LPG) Activates NK Cells Through Toll-Like Receptor-2. *Mol Biochem Parasitol* (2003) 130:65–74. doi: 10.1016/S0166-6851(03)00160-9
57. Messlinger H, Sebald H, Heger L, Dudziak D, Bogdan C, Schleicher U. Monocyte-Derived Signals Activate Human Natural Killer Cells in Response to Leishmania Parasites. *Front Immunol* (2018) 9:24. doi: 10.3389/fimmu.2018.00024
58. Arun G, Aggarwal D, Spector DL. MALAT1 Long Non-Coding RNA: Functional Implications. *Non-Coding RNA* (2020) 6:22. doi: 10.3390/ncrna6020022
59. Chen Q, Zhu C, Jin Y. The Oncogenic and Tumor Suppressive Functions of the Long Noncoding RNA MALAT1: An Emerging Controversy. *Front Genet* (2020) 11:93. doi: 10.3389/fgene.2020.00093
60. Hewitson JP, West KA, James KR, Rani GF, Dey N, Romano A, et al. Malat1 Suppresses Immunity to Infection Through Promoting Expression of Maf and IL-10 in Th Cells. *J Immunol* (2020) 204:2949–60. doi: 10.4049/jimmunol.1900940
61. Wu J, Zhang H, Zheng Y, Jin X, Liu M, Li S, et al. The Long Noncoding RNA MALAT1 Induces Tolerogenic Dendritic Cells and Regulatory T Cells *via* Mir155/Dendritic Cell-Specific Intercellular Adhesion Molecule-3 Grabbing Nonintegrin/IL10 Axis. *Front Immunol* (2018) 9:1847. doi: 10.3389/fimmu.2018.01847
62. Ulitsky I, Bartel DP. lincRNAs: Genomics, Evolution, and Mechanisms. *Cell* (2013) 154:26–46. doi: 10.1016/j.cell.2013.06.020
63. Gil N, Ulitsky I. Regulation of Gene Expression by Cis -Acting Long Non-Coding RNAs. *Nat Rev Genet* (2020) 21:102–17. doi: 10.1038/s41576-019-0184-5
64. Tan JY, Smith AAT, Ferreira da Silva M, Matthey-Doret C, Rueedi R, Sönmez R, et al. Cis-Acting Complex-Trait-Associated lincRNA Expression Correlates With Modulation of Chromosomal Architecture. *Cell Rep* (2017) 18:2280–8. doi: 10.1016/j.celrep.2017.02.009
65. Huang W, Liu Z, Li Y, Liu L, Mai G. Identification of Long Noncoding RNAs Biomarkers for Diagnosis and Prognosis in Patients With Colon Adenocarcinoma. *J Cell Biochem* (2019) 120:4121–31. doi: 10.1002/jcb.27697
66. Zhang X-Y, Zhuang H-W, Wang J, Shen Y, Bu Y-Z, Guan B-G, et al. Long Noncoding RNA CA3-AS1 Suppresses Gastric Cancer Migration and Invasion by Sponging miR-93-5p and Targeting BTG3. *Gene Ther* (2020) 1–9. doi: 10.1038/s41434-020-00201-1
67. Wei H, Yang Z, Lin B. Overexpression of Long Non Coding RNA CA3-AS1 Suppresses Proliferation, Invasion and Promotes Apoptosis *via* miRNA-93/PTEN Axis in Colorectal Cancer. *Gene* (2019) 687:9–15. doi: 10.1016/j.gene.2018.11.008
68. Tay Y, Rinn J, Pandolfi PP. The Multilayered Complexity of ceRNA Crosstalk and Competition. *Nature* (2014) 505:344–52. doi: 10.1038/nature12986
69. Deng H, Ouyang W, Zhang L, Xiao X, Huang Z, Zhu W. LncRNA GASL1 Is Downregulated in Chronic Heart Failure and Regulates Cardiomyocyte Apoptosis. *Cell Mol Biol Lett* (2019) 24:41. doi: 10.1186/s11658-019-0165-x
70. Su W-Z, Yuan X. LncRNA GASL1 Inhibits Tumor Growth of Non-Small Cell Lung Cancer by Inactivating TGF-β Pathway. *Eur Rev Med Pharmacol Sci* (2018) 22:7282–8. doi: 10.26355/eurrev_201811_16264
71. Shi B, Shao B, Yang C, Guo Y, Fu X, Gan N. Upregulation of JHDM1D-AS1 Protects PDLSCs From H2O2-Induced Apoptosis by Decreasing DNAJC10 *via* Phosphorylation of Eif2α. *Biochimie* (2019) 165:48–56. doi: 10.1016/j.biochi.2019.06.018
72. Qiu B-Q, Lin X-H, Ye X-D, Huang W, Pei X, Xiong D, et al. Long Non-Coding RNA PSMA3-AS1 Promotes Malignant Phenotypes of Esophageal Cancer by Modulating the miR-101/EZH2 Axis as a ceRNA. *Aging (Albany NY)* (2020) 12:1843–56. doi: 10.18632/aging.102716
73. Wang L, Wu L, Pang J. Long Noncoding RNA PSMA3−AS1 Functions as a microRNA−409−3p Sponge to Promote the Progression of Non−Small Cell Lung Carcinoma by Targeting Spindlin 1. *Oncol Rep* (2020) 44:1550–60. doi: 10.3892/or.2020.7693
74. Kent LN, Leone G. The Broken Cycle: E2F Dysfunction in Cancer. *Nat Rev Cancer* (2019) 19:326–38. doi: 10.1038/s41568-019-0143-7
75. Nascimento MSL, Ferreira MD, Quirino GFS, Maruyama SR, Krishnaswamy JK, Liu D, et al. NOD2-RIP2–Mediated Signaling Helps Shape Adaptive Immunity in Visceral Leishmaniasis. *J Infect Dis* (2016) 214:1647–57. doi: 10.1093/infdis/jiw446
76. Murphy TL, Tussiwand R, Murphy KM. Specificity Through Cooperation: BATF-IRF Interactions Control Immune-Regulatory Networks. *Nat Rev Immunol* (2013) 13:499–509. doi: 10.1038/nri3470
77. Liu Q, Kim MH, Friesen L, Kim CH. BATF Regulates Innate Lymphoid Cell Hematopoiesis and Homeostasis. *Sci Immunol* (2020) 5:eaaz8154. doi: 10.1126/sciimmunol.aaz8154
78. Panda SK, Colonna M. Innate Lymphoid Cells in Mucosal Immunity. *Front Immunol* (2019) 10:861. doi: 10.3389/fimmu.2019.00861
79. Xu X, Gu J, Ding X, Ge G, Zang X, Ji R, et al. LINC00978 Promotes the Progression of Hepatocellular Carcinoma by Regulating EZH2-Mediated Silencing of P21 and E-Cadherin Expression. *Cell Death Dis* (2019) 10:752. doi: 10.1038/s41419-019-1990-6
80. Zhang S, Li C, Liu J, Geng F, Shi X, Li Q, et al. Fusobacterium Nucleatum Promotes Epithelial-Mesenchymal Transiton Through Regulation of the lncRNA MIR4435-2hg/miR-296-5p/Akt2/SNAI1 Signaling Pathway. *FEBS J* (2020) 287:4032–47. doi: 10.1111/febs.15233
81. Patamsytė V, Žukovas G, Gečys D, Žaliaduonytė D, Jakuška P, Benetis R, et al. Long Noncoding RNAs CARMN, LUCAT1, SMILR, and MALAT1 in Thoracic Aortic Aneurysm: Validation of Biomarkers in Clinical Samples. *Dis Markers* (2020) 2020:8521899. doi: 10.1155/2020/8521899
82. Gong J, Xu X, Zhang X, Zhou Y. LncRNA MIR4435-2HG Is a Potential Early Diagnostic Marker for Ovarian Carcinoma. *Acta Biochim Biophys Sin* (2019) 51:953–9. doi: 10.1093/abbs/gmz085
83. Xu H, Han H, Song S, Yi N, Qian C, Qiu Y, et al. Exosome-Transmitted PSMA3 and PSMA3-AS1 Promote Proteasome Inhibitor Resistance in Multiple Myeloma. *Clin Cancer Res* (2019) 25:1923–35. doi: 10.1158/1078-0432.CCR-18-2363
84. Wang Y-M, Trinh MP, Zheng Y, Guo K, Jimenez LA, Zhong W. Analysis of Circulating Non-Coding RNAs in a Non-Invasive and Cost-Effective Manner. *TrAC Trends Analyt Chem* (2019) 117:242–62. doi: 10.1016/j.trac.2019.07.001
85. Li Q, Shao Y, Zhang X, Zheng T, Miao M, Qin L, et al. Plasma Long Noncoding RNA Protected by Exosomes as a Potential Stable Biomarker for Gastric Cancer. *Tumour Biol* (2015) 36:2007–12. doi: 10.1007/s13277-014-2807-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Immunology

ORIGINAL RESEARCH
published: 17 March 2022
doi: 10.3389/fimmu.2022.836533

Osteoarticular Involvement-Associated Biomarkers and Pathways in Psoriasis: The Shared Pathway With Ankylosing Spondylitis

Yu-Ping Zhang[†], *Xing Wang*[†], *Li-Gang Jie*, *Yuan Qu*, *Xiao-Tong Zhu*, *Jing Wu*[*] *and Qing-Hong Yu*[*]

Department of Rheumatology and Clinical Immunology, Zhujiang Hospital, Southern Medical University, Guangzhou, China

OPEN ACCESS

Edited by:
Xu-jie Zhou,
Peking University First Hospital, China

Reviewed by:
Magdalena Krajewska-Włodarczyk,
University of Warmia and Mazury in Olsztyn, Poland
Xiaoxiang Chen,
Shanghai JiaoTong University, China

***Correspondence:**
Jing Wu
wujing211@163.com
Qing-Hong Yu
yuqinghong@smu.edu.cn

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Autoimmune and Autoinflammatory Disorders, a section of the journal Frontiers in Immunology

Received: *15 December 2021*
Accepted: *18 February 2022*
Published: *17 March 2022*

Citation:
Zhang Y-P, Wang X, Jie L-G, Qu Y, Zhu X-T, Wu J and Yu Q-H (2022) Osteoarticular Involvement-Associated Biomarkers and Pathways in Psoriasis: The Shared Pathway With Ankylosing Spondylitis. Front. Immunol. 13:836533. doi: 10.3389/fimmu.2022.836533

Psoriatic arthritis (PsA) is a unique immune-mediated disease with cutaneous and osteoarticular involvement. However, only a few studies have explored the susceptibility of osteoarticular involvement in psoriasis (Ps) at the genetic level. This study investigated the biomarkers associated with osteoarticular participation and potential shared molecular mechanisms for PsA and ankylosing spondylitis (AS).

Methods: The RNA-seq data of Ps, PsA, and AS in the Gene Expression Omnibus (GEO) database were obtained. First, we used the limma package and the weighted gene co-expression network analysis (WGCNA) to identify the potential genes related to PsA and AS. Then, the shared genes in PsA and AS were performed using the GO, KEGG, and GSEA analyses. We also used machine learning to screen hub genes. The results were validated using external datasets and native cohorts. Finally, we used the CIBERSORT algorithm to estimate the correlation between hub genes and the abundance of immune cells in tissues.

Results: An overlap was observed between the PsA and AS-related modules as 9 genes. For differentially expressed genes in AS and PsA, only one overlapping gene was found (COX7B). Gene enrichment analysis showed that the above 9 genes might be related to the mRNA surveillance pathway. The GSEA analyses showed that COX7B was involved in adaptive immune response, cell activation, etc. The PUM1 and ZFP91, identified from the support vector machine, had preferable values as diagnostic markers for osteoarticular involvement in Ps and AS (AUC > 0.7). Finally, CIBERSORT results showed PUM1 and ZFP91 involvement in changes of the immune microenvironment.

Conclusion: For the first time, this study showed that the osteoarticular involvement in psoriasis and AS could be mediated by the mRNA surveillance pathway-mediated abnormal immunologic process. The biological processes may represent the cross talk between PsA and AS. Therefore, PUM1 and ZFP91 could be used as potential biomarkers or therapeutic targets for AS and Ps patients.

Keywords: psoriasis, psoriatic arthritis, ankylosing spondylitis, WGCNA, differential gene analysis

INTRODUCTION

Psoriasis (Ps) is a chronic inflammatory skin disease with an incidence rate of 1% to 3%. Arthritis occurring in 10%–40% of psoriasis patients is called psoriatic arthritis (PsA) (1). In most patients with PsA, skin manifestations appear first, preceding arthritis over several years (2). Terminal stages of PsA are generally characterized by joint deformity and/or spinal ankylosis. Osteoarticular involvement represents any symptoms and signs of osteoarticular system, including spondylitis, enthesitis, peripheral arthritis, and dactylitis which affects the patient's overall quality of life. Previous studies have compared the differences in the genetic background between psoriasis and PsA (3–5). However, as genomically similar but phenotypically distinct diseases, more insight is required at the transcriptomic level to understand the biomarkers or biological pathways associated with the development of osteoarticular involvement in Ps and AS (6).

PsA is a member of the spondyloarthropathy (SpA) family. The SpA diseases, including psoriatic arthritis, ankylosing spondylitis (AS), arthritis associated with inflammatory bowel disease (IBD), reactive arthritis, and undifferentiated spondyloarthropathy, share common genetic backgrounds and present overlapping clinical signs. The ankylosing spondylitis is the prototype of the SpA group. In addition, the bioinformatics background of AS can be a potential representative of osteoarticular involvement, especially axial involvement in spondyloarthropathy (7).

Therefore, this study was carried out to explore osteoarticular involvement-associated biomarkers and pathways in psoriasis patients. We used the limma package and the weighted gene co-expression network analysis (WGCNA) to identify the potential common genes related to PsA and AS. Additionally, we analyzed the published gene expression data at the Gene Expression Omnibus (GEO). We showed that the osteoarticular involvement in psoriasis and AS could be associated with mRNA surveillance pathway-mediated abnormal processes. Moreover, the identified PUM1 and ZFP91 genes were preferable as diagnostic markers for osteoarticular involvement in Ps and AS. To the best of our knowledge, this is the first study to use a systemic bioinformatic analysis approach to explore the gene signatures of osteoarticular involvement in Ps and AS.

MATERIALS AND METHODS

Datasets and Data Preprocessing

We obtained the GEO database's original gene expression profile data and clinical information (8). We used the keywords "psoriatic arthritis" and "psoriasis" or "ankylosing spondylitis" to search RNA-seq profiles in the GEO database. The following filter criteria were used: the organization used for sequencing should be peripheral blood mononuclear cells (PBMC), and the number of samples of each group should not be less than 10 to ensure the accuracy of the WGCNA. Finally, the GEO datasets numbered GSE61281, GSE25101, and GSE73754 were obtained.

The GSE61281 dataset was used on the GPL6480 platform. The dataset contained 40 samples, including peripheral blood samples from cutaneous psoriasis without inflammatory arthritis (n = 20) and 20 peripheral blood samples from PsA (n = 20). The GSE25101 dataset was used on the GLP6947 platform. This dataset contained 32 samples, including 16 peripheral blood samples from AS patients and 16 peripheral blood samples from healthy controls. Besides, the GSE73754 dataset based on GPL10558 was regarded as an external validation set from the GEO database with 51 cases of AS patients as the experimental group and 20 cases of normal samples as the control group. Moreover, to estimate the diagnostic efficiency of skin lesions other than blood biomarkers, the GSE13355 dataset based on GPL570 was downloaded, including 58 psoriatic lesional samples and 64 non-psoriasis skin samples. The detailed clinical characteristics are shown in **Supplementary Files 1**.

Screening and Validation of Hub Biomarkers

Firstly, we utilized the limma R package to screen the differential genes (DEGs) from GSE61281 and GSE25101 datasets (9). The screening conditions for the DEGs were the absolute value of |log2 fold change FC|<0.5, and adj. p-value <0.05 was considered as the standard. Next, WGCNA was performed on the obtained DEGs from two datasets. Based on the scale-free topology criterion, the soft-power parameters ranging from 1 to 20 using the "pickSoftThreshold" (package WGCNA) function were screened out. The extracted values were chosen to build an adjacency matrix (10). The most appropriate β value was selected to convert the matrix of correlations to the adjacency matrix and then into a topological overlap matrix. Next, we used the average-linkage hierarchical clustering method to cluster genes based on TOM, where the minimum module size was set at 50. After that, modules with similarities were merged. Finally, Pearson correlation analysis was performed to assess the correlation of the integrated modules with the osteoarticular involvement in Ps and AS.

The support vector machine-recursive feature elimination (SVM-RFE) (11) is a sequential backward feature elimination method based on SVM, which is used to find the optimal hub gene by deleting feature vectors dependent on the e1071 and msvmRFE package (12)) for SVM modeling. We screened core biomarkers using SVM analysis in the above DEGs and intersection of WCGNA. The area under the ROC curve (AUC) was evaluated to assess the diagnostic performance of the core biomarker on the datasets (GSE73754 and GSE13355) and the sequencing data of samples from our hospital.

Enrichment Analysis

Gene Ontology (GO) category analysis is commonly used for the bioinformatics analysis of large datasets (13). Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database resource for understanding the high-level functions and utilities of the biological system. The results from the GO and KEGG analyses were visualized using the "GOplot" package in R software. Finally, the cluster profile and GSVA packages were used to explore the COX7B gene correlation with specific signaling pathways (14, 15). The gene sets were downloaded from MSigDB (c5.go.bp.v7.4.symbols.gmt) (16). The potential pathways of gene sets and gene expression matrix were detected using gene set enrichment analysis (GSEA).

Construction of Hub Gene Regulatory Network

The TF–gene interaction pairs with p-values <0.05 were retrieved from TRRUST (17). Finally, the visualization of core gene regulatory network was implemented by using NetworkAnalyst (18).

Immune Analysis Algorithm

CIBERSORT is a deconvolution algorithm that combines the labeled genomes of different immune cell subpopulations to calculate the proportion of 22 immune cells in tissues (19). Non-parametric correlations (Spearman) were used to determine the correlation between core biomarkers and immune-infiltrated cells.

RESULTS

Differential Gene Screening

Based on the GSE61281 dataset, a total of 37 differential genes (DEGs) were identified. The heatmap demonstrates the top 10 DEGs (**Figure 1A**), obtained using the logFC value. The volcano plot shows the identified DEGs, including 25 upregulated and 12 downregulated (**Figure 1C**). Besides, a total of 62 DEGs were obtained from the GSE25101 dataset, among which 42 genes were upregulated and 20 were found to be downregulated (**Figure 1D**). Heatmaps of the top 10 upregulated and downregulated DEGs are shown in **Figure 1B**.

Weighted Gene Co-Expression Network Analysis

We performed WGCNA to investigate the correlation between the clinical information and key genes. The genes with significant differential expression ($p < 0.05$) were selected as inputs of WGCNA. All samples were clustered in the GSE61281 and

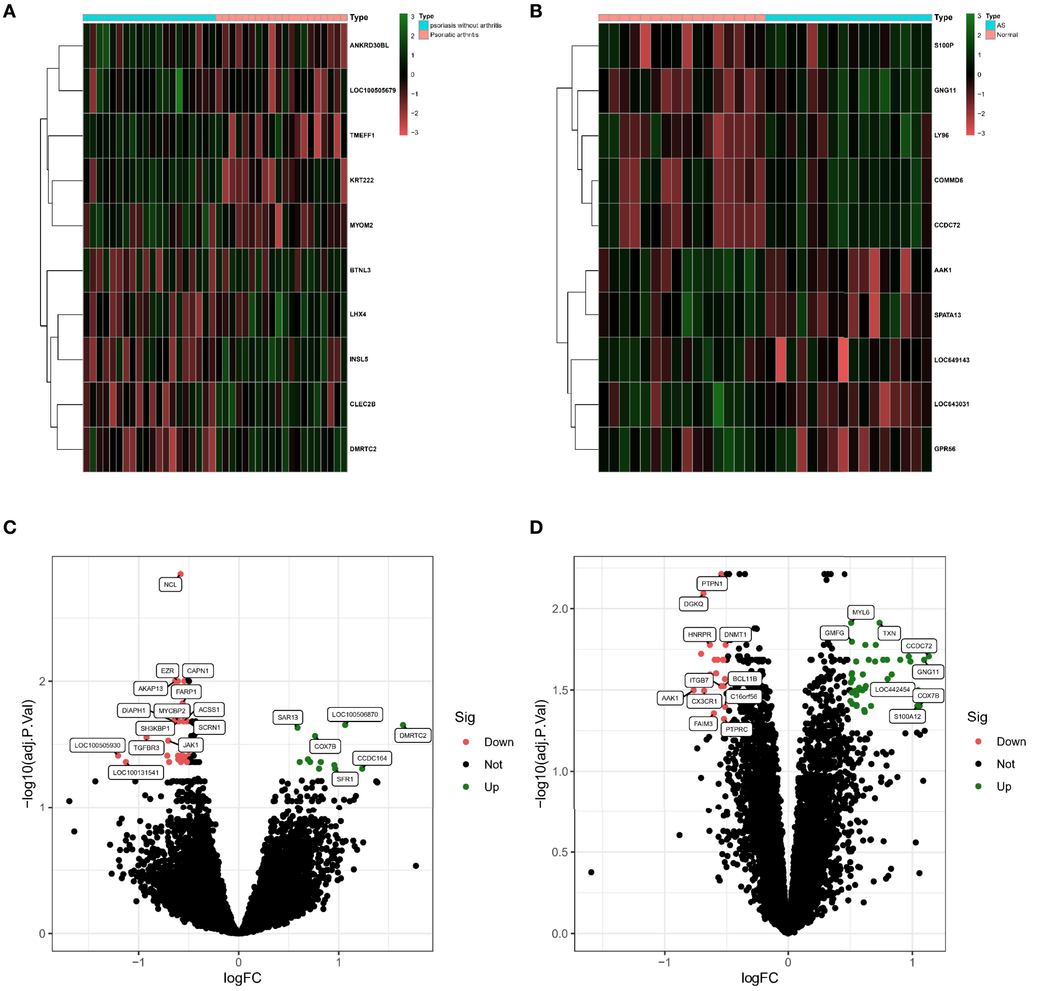


FIGURE 1 | **(A)** Heatmap of DEGs in GSE61281 (n = 37, adj. p < 0.05, |log2 fold change FC| < 0.5). **(B)** Heatmap of DEGs in GSE25101 (n = 62, adj. p < 0.05, |log2 fold change FC| > 0.5). **(C)** Volcano plot of DEGs in GSE61281. **(D)** Volcano plot of DEGs in GSE25101.

GSE25101 datasets, and none of the samples was eliminated (**Figures 2A, B**). In the WGCNA methodology, β = 6 was the optimal soft-power value for GSE61281 (**Figure 2C**), and β = 11 was the optimal soft-power value for GSE25101 (**Figure 2D**). A total of 13 modules were identified in GSE61281, and 6 were identified in GSE25101. Afterward, the correlations between the module and clinical traits were calculated. The gray module had the strongest positive relation with PsA (r = 0.6), while the pink module had the strongest negative relation (r = 0.64) in the GSE61281 database (**Figure 2E**). For AS, the gray module showed the strongest positive correlation (r = 0.9), and cyan had the strongest negative correlation (r = -0.68) in the GSE25101 database (**Figure 2F**).

Identification of the Shared Genes and TF-mRNA Regulatory Network

An overlap was observed between the PsA and AS modules as a total of 9 genes (TTC3, ZFP91, MACF1, BDP1, PUM1, SRRM1, SUPT16H, PABPC3, ZNF135) (**Figure 3A**). For DEGs, only one overlapping gene was found (COX7B) (**Figure 3B**). These genes might be involved in the pathogenesis of osteoarticular involvement in psoriasis and AS, and have a sharing relationship. Therefore, we searched for an upstream transcriptional regulator that possibly regulated the above 10 genes by the JASPAR database based on the above results. There were 46 nodes and 67 edges found in total (**Figure 3C**).

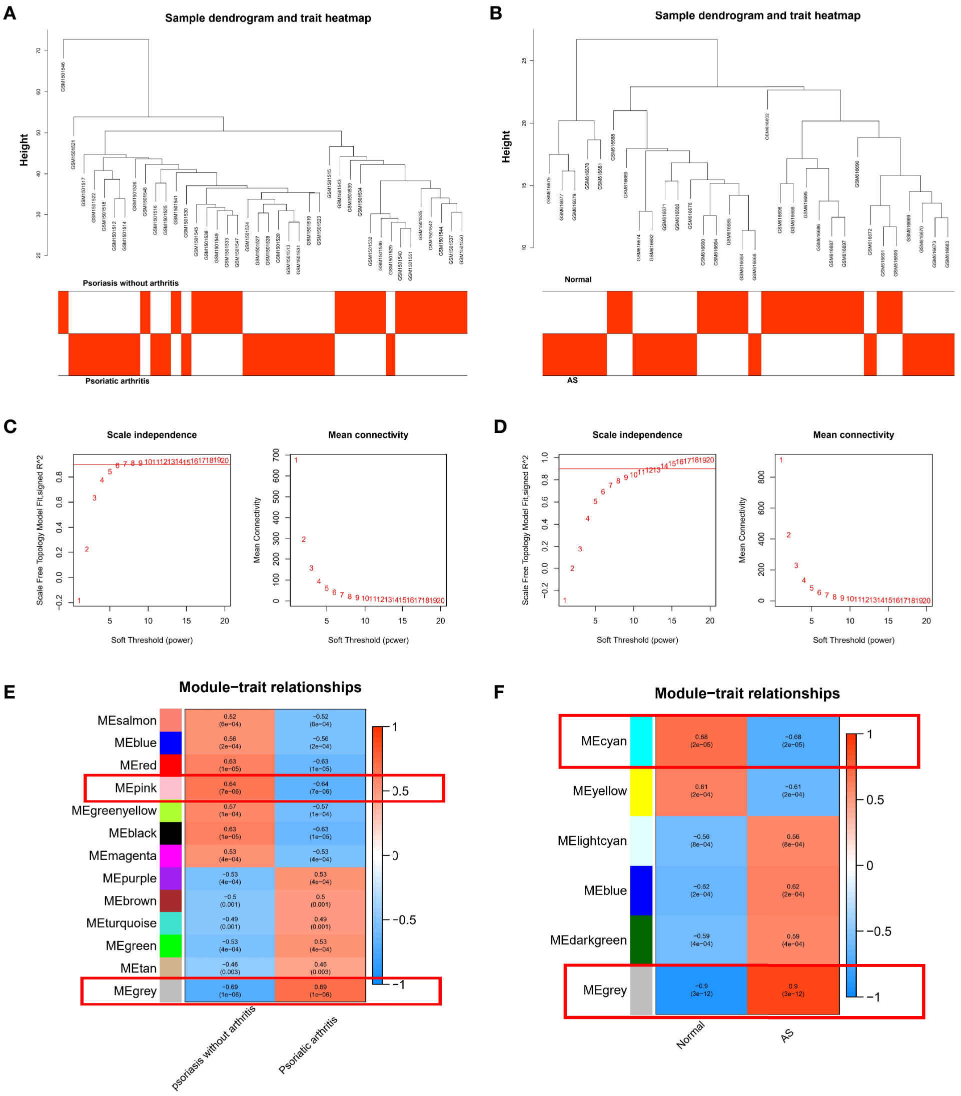


FIGURE 2 | **(A)** Correlation between modules and genes in GSE61281. **(B)** Correlation between modules and genes in GSE25101. **(C)** Determination of soft-thresholding power for GSE61281. **(D)** Determination of soft-thresholding power for GSE25101**. (E)** Heatmap of the correlation between module eigengenes and the occurrence of PsA. **(F)** Heatmap of the correlation between module eigengenes and the occurrence of AS.

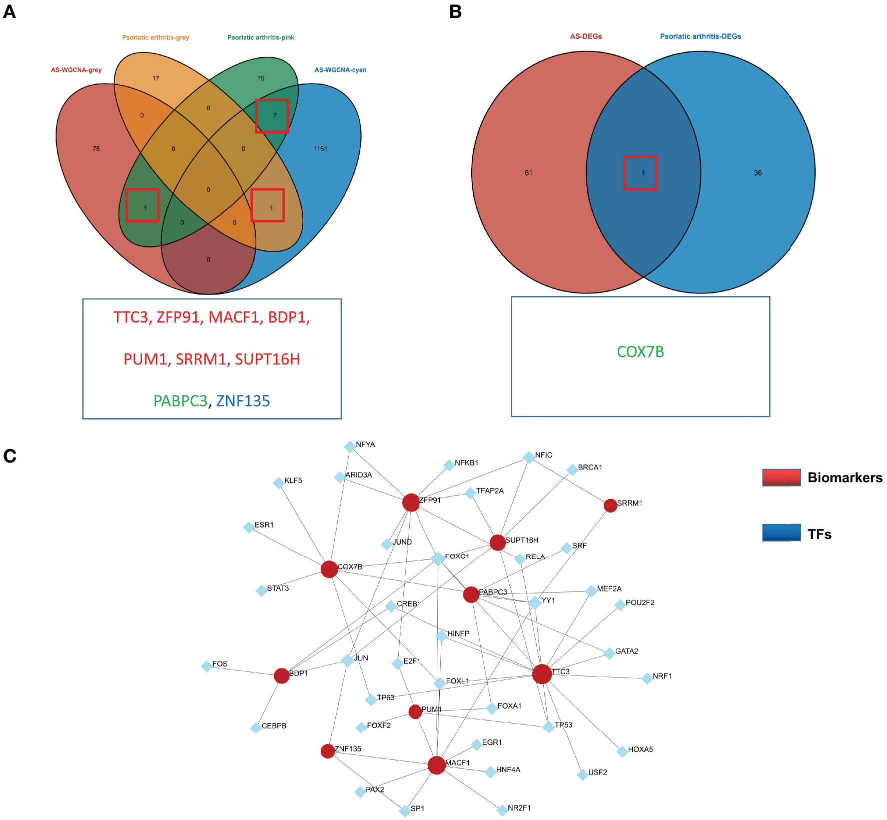


FIGURE 3 | **(A)** Venn diagram shows an overlap of 9 genes in modules between PsA and AS. **(B)** Venn diagram shows an overlap of one DEGs between PsA and AS. **(C)** TF-miRNA regulatory networks. Blue nodes represent transcription factors (TFs), and red nodes represent biomarkers. Black edges represent regulatory relationships between TFs and biomarkers.

Identification of the Shared Pathways

We further explored the common regulatory pathway that 9 genes were screened by WGCNA and 1 gene was selected as overlapping DEGs. The GO and KEGG enrichment analyses were performed in the above 9 genes. The GO analysis showed that the above 9 genes might be related to the cytoplasmic stress granule, cytoplasmic ribonucleoprotein granule, ribonucleoprotein granule, and RNA polymerase III transcription factor complex (**Figure 4A**). The KEGG analysis showed that these genes might be correlated with the mRNA surveillance pathway (**Figure 4B**). Finally, we performed a single-gene GSEA analysis. The only shared differential gene of AS and PsA samples (COX7B) might participate in several biological processes, including adaptive immune response and cell activation (**Figures 4C, D**). Hence, we made a conjecture that the occurrence of osteoarticular involvement in psoriasis and AS was likely mediated by mRNA surveillance pathway-mediated abnormal immunologic processes.

Identification of Potential Shared Diagnostic Gene Targets Based on the Machine Learning Algorithm

SVM-RFE is a machine learning method based on the support vector machines used to find the best core gene by deleting feature vectors produced by SVM. Based on the 10 shared genes, a total of 8 genes were identified as the biomarkers in GSE25101 (**Figure 5A**), and 2 genes in the GSE61281 dataset (**Figure 5B**). These biomarkers might have diagnostic value. Finally, we identified PUM1 and ZFP91 as the optimal diagnostic biomarkers for osteoarticular involvement in psoriasis and AS (**Figure 5C**).

Validation of Diagnostic Shared Biomarkers

Furthermore, we identified the diagnostic efficacy of the shared biomarkers. In the GSE25101 dataset, these two biomarkers had preferable values as diagnostic markers: PUM1 (AUC = 0.733) and ZFP91 (AUC = 0.836) (**Figure 6A**). The same ROC analysis was performed again for the above biomarkers in the GSE61281 dataset. Each biomarker showed the robust capacity of predictive performance: PUM1 (AUC = 0.970) and ZFP91 (AUC = 0.872) (**Figure 6B**). We then performed external validation for the diagnostic efficacy of PUM1 and ZFP91 in GSE73754, similar to blood RNA-sequencing in AS. The results showed that they show significant differences in expression (**Figure 6C**) and good diagnostic accuracy for detection of AS: PUM1 (AUC = 0.638) and ZFP91 (AUC = 0.642) (**Figure 6E**). In data of samples from our hospital, only PUM1 (AUC = 0.0.889) showed good prediction

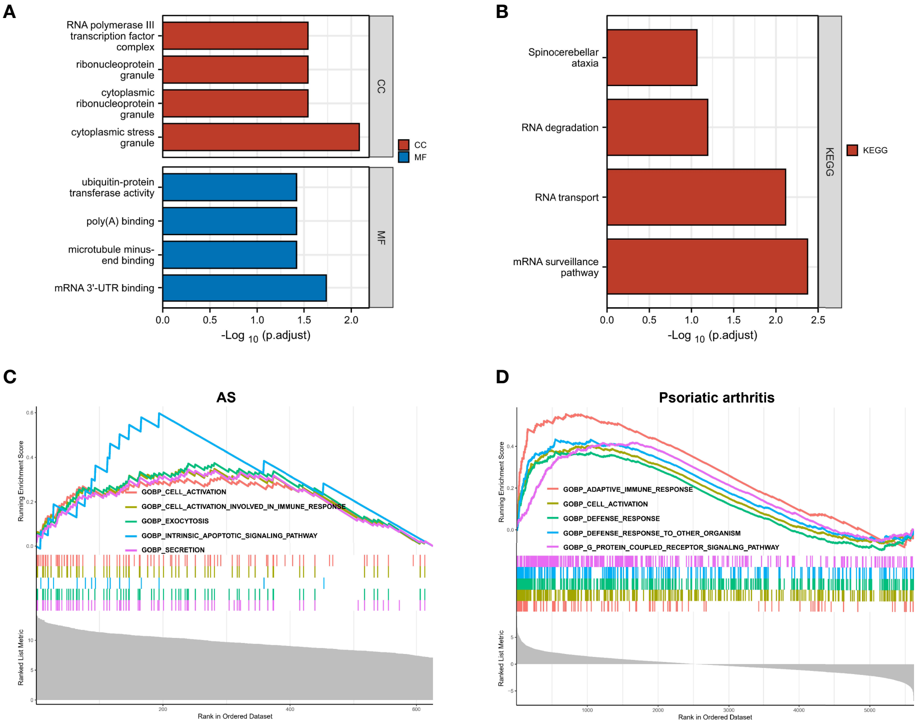


FIGURE 4 | **(A)** GO enrichment analysis results for 9 genes screened by WGCNA. **(B)** KEGG enrichment analysis results for 9 genes screened by WGCNA. **(C)** Single-gene GSEA analyses results for DEGs of AS. **(D)** Single-gene GSEA analyses results for DEGs of PsA.

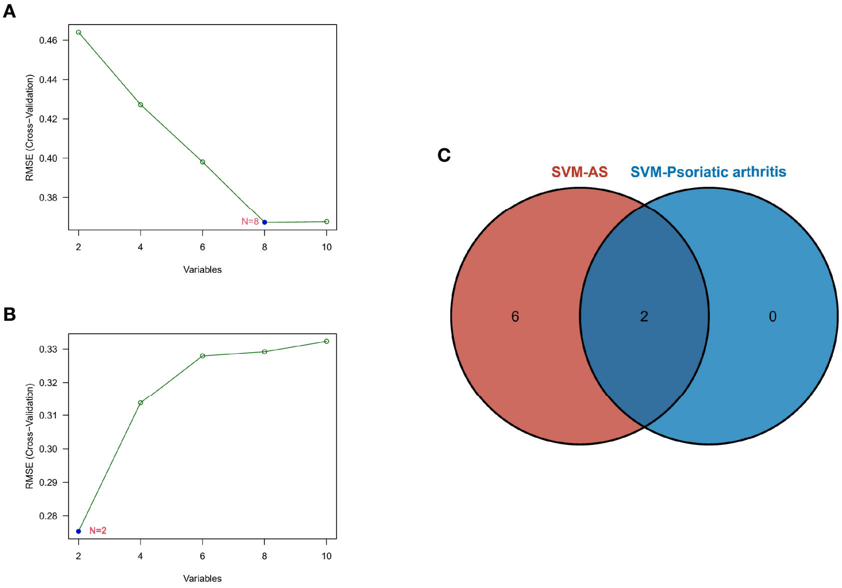


FIGURE 5 | **(A)** SVM-RFE algorithm to screen diagnostic markers in the GSE25101 database. **(B)** SVM-RFE algorithm to screen diagnostic markers in the GSE61281 database. **(C)** Venn diagram shows the optimal diagnostic biomarkers.

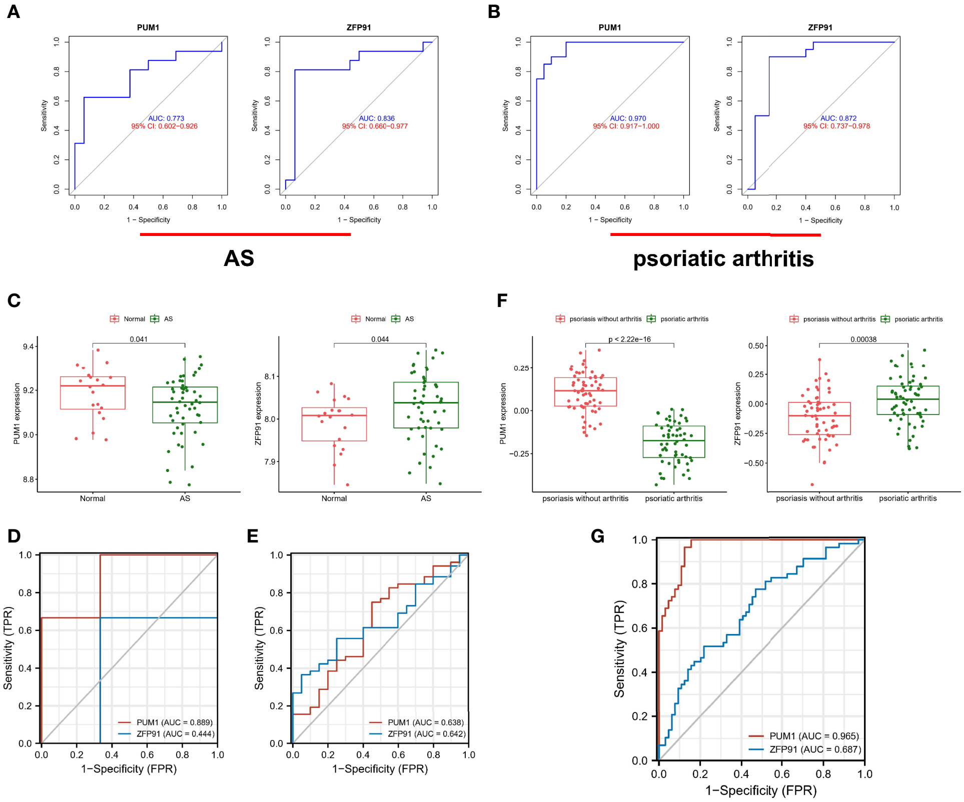


FIGURE 6 | Validation of diagnostic shared biomarkers. **(A)** The ROC curve of the diagnostic efficacy verification in GSE25101. **(B)** The ROC curve of the diagnostic efficacy verification in GSE61281. **(C)** The shared biomarkers in GSE73754 showed significant differences, with p value < 0.05. **(D)** The ROC curve of the diagnostic efficacy verification in data of samples from our hospital. **(E)** The ROC curve of the diagnostic efficacy verification in GSE73754. **(F)** The shared biomarkers in GSE13355 showed significant differences, with p value < 0.05. **(G)** The ROC curve of the diagnostic efficacy verification in GSE13355.

efficacy, while ZFP91 showed weak prediction efficacy (**Figure 6D**). Meanwhile, we analyzed RNA-seq datasets of skin samples of psoriasis (GSE13355). Both PUM1 and ZFP91 showed significant differences between groups (**Figure 6F**). Each biomarker had potent predictive performance: PUMI (AUC = 0.965) and ZFP91 (AUC = 0.687) (**Figure 6G**).

Immune Infiltration Analysis of Shared Biomarkers

The enrichment analysis results showed that immunity plays an important role in developing osteoarticular involvement in psoriasis and AS. The CIBERSORT algorithm was used to analyze the abundances of immune cells in different samples. Bar graphs show the significant differences in the percentage of B cell and macrophage populations between AS and psoriasis samples (**Figures 7A**, **8A**). Compared with the normal sample, CD4-naive T cells and regulatory T cells (Tregs) were decreased in the AS sample, while monocytes were increased (**Figure 7B**). However, compared with psoriasis without arthritis, T cells CD4 memory activated were increased in the psoriasis with osteoarticular involvement, while T cells CD4 memory resting decreased (**Figure 8B**). Moreover, the correlation of the biomarkers and content of different immune cells was explored. In AS samples, PUM1 had a significant positive correlation with naive B cells and CD4-naive T cell content. In contrast, there was a significant negative correlation between PUM1 and both monocytes and activated dendritic cell content (**Figure 7C**). ZFP91 had a significant positive correlation with both CD8 T cells and Tregs and a significant negative correlation with monocytes, activated dendritic cells, and neutrophils (**Figure 7D**). In PsA samples, only PUM1 had a significant positive correlation with memory resting CD4T cells (**Figure 8C**). The statistical analysis showed no significant differences between ZFP91 and other immune cell content (**Figure 8D**).

DISCUSSION

The characteristics and phenotype of osteoarticular involvement of PsA is consistent with the phenotype of spondyloarthritis (20–22). Meanwhile, ankylosing spondylitis is the prototype of spondyloarthritis, with the typical characteristics of osteoarticular involvement of SpA, including spondylitis, enthesitis, peripheral arthritis, and dactylitis. Osteoarticular involvement of SpA may have a variety of manifestations, but the pathological mechanism seems to be similar, such as pathologic new bone formation,

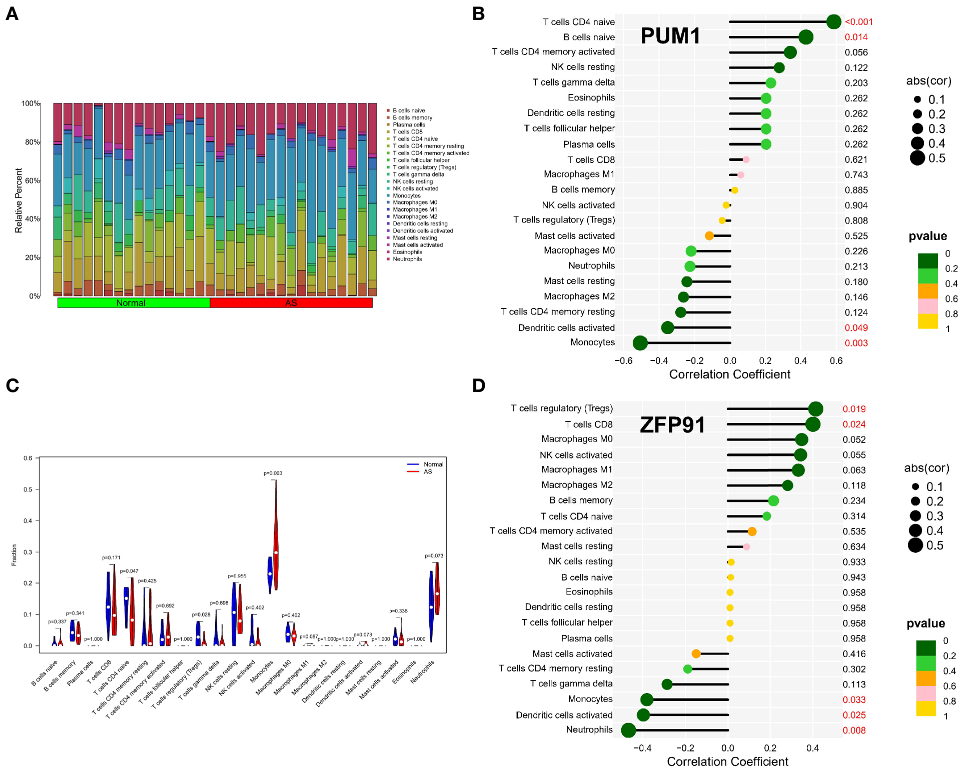


FIGURE 7 | Immune infiltration analysis of shared biomarkers in AS**. (A)** The barplot of immune cell infiltration. **(B)** Correlation between PUM1 and infiltrating immune cells. **(C)** Violin diagram of the proportion of 22 types of immune cells. The red marks represent the difference in infiltration between the two groups of samples. **(D)** Correlation between ZFP91 and infiltrating immune cells.

typically occurring at sites of soft tissue surrounding the entheses, and synovitis with more vascularity, a greater infiltration of neutrophils than rheumatic arthritis (22). Therefore, considering the phenotypic disturbances in the PsA group, we studied osteoarticular involvement as a clinical feature that differentiated PsA from Ps. The WGCNA approach was successfully applied in various diseases to identify common risk biomarkers and pathways associated with the different disease phenotypes (23–25). In the present study, the shared gene co-expression module of AS and PsA revealed that the osteoarticular involvement in Ps and AS could be associated with the mRNA surveillance pathway-mediated abnormal immunologic process and identified PUM1 and ZFP91 had preferable value as diagnostic markers for osteoarticular involvement, especially axial involvement in Ps and AS.

The mRNA surveillance pathway is a quality-control physiological mechanism that degrades and detects abnormal mRNAs, including non-stop mRNA decay (NSD), nonsense-mediated mRNA decay (NMD), and no-go decay (NGD) (**Supplementary Figure S1**). The present study results show that the overlapping genes between the PsA and AS modules in WGCNA were correlated with the mRNA surveillance pathway. This observation was suggestive of the pathophysiologic mechanism of osteoarticular involvement in Ps association with the mRNA surveillance pathway. PsA is a chronic immune-mediated rheumatic disease. Studies have reported that the posttranscriptional regulation of gene expression plays a vital role in rheumatic disease (26, 27). However, most studies have focused on microRNA and related pathways, while only a few focused on the mRNA surveillance pathway (28, 29). Notably, no studies have been conducted for assessing the relationship of osteoarticular involvement in Ps and mRNA surveillance pathway; thus, the specific mechanism needs to be further studied and confirmed. Moreover, multiple factors are involved in the immunologic process and mRNA surveillance pathway; the results of this study represent the preliminary exploration.

This pathway was not reported for Ps or PsA, but it was confirmed to be involved in the mechanism of other immune-mediated disease. The abnormal mRNA surveillance machinery causes abnormal activation of immunologic defense programs, resulting in autoimmune diseases (30). As an immune-mediated disease, the process of osteoarticular involvement on psoriasis is probably associated with this mechanism. However, considering the complexity of the mRNA surveillance pathway, the specific mechanism still needs to be confirmed by further studies.

Meanwhile, the COX7B gene is shared between the DEGs of PsA and AS. Single-gene GSEA analyses suggested that it might be associated with several immunologic processes, including adaptive immune response and cell activation. Moreover, the immunologic process and inflammatory response of SpA including PsA are different from other inflammatory

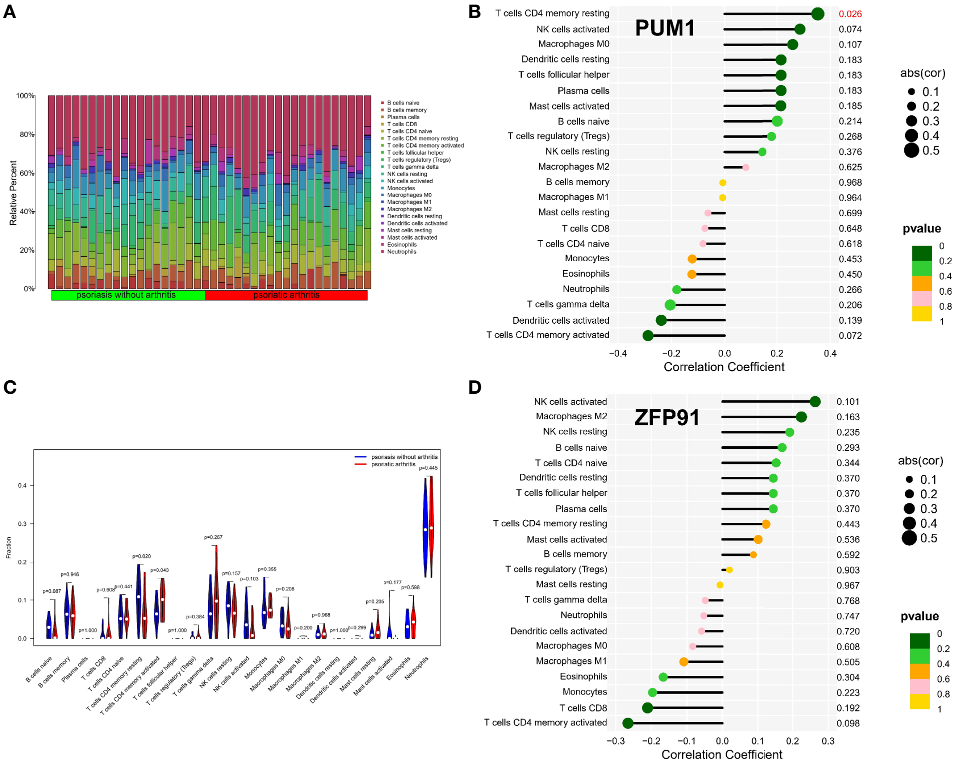


FIGURE 8 | Immune infiltration analysis of shared biomarkers in PsA**. (A)** The barplot of immune cell infiltration. **(B)** Correlation between PUM1 and infiltrating immune cells. **(C)** Violin diagram of the proportion of 22 types of immune cells. The red marks represent the difference in infiltration between the two groups of samples. **(D)** Correlation between ZFP91 and infiltrating immune cells.

arthritis (31). However, concerning immune response and epigenetics, Ps and PsA were usually considered as a group of diseases in most studies; therefore, we barely knew the difference of immunologic processes between psoriasis and PsA (32). The main reason probably lies in the challenge to obtain samples of involved joints. In this case, the analysis of the RNA-seq profiles of PBMCs is an alternative. Our study showed differences in the Ps- and PsA-immune microenvironment, correlating with the biomarkers of osteoarticular involvement.

Immune infiltration analysis revealed the involvement of several specialized immune cell populations, suggesting differences in the immune response between PsA and Ps. Compared with psoriasis without arthritis, T cells CD4 memory activated were increased in psoriasis with osteoarticular involvement, while T cells CD4 memory resting were decreased. Similarly, recent studies have demonstrated that tissue-resident memory CD8+ T cells from the skin helped differentiate psoriatic arthritis from psoriasis (33). Of note, CD4+T cells play an important role in PsA (34).

Furthermore, our studies have shown that PUM1 and ZFP91 might be useful biomarkers or potential therapeutic targets for osteoarticular involvement in Ps and AS due to their involvement in the pathophysiology of osteoarticular involvement of PsA. Previously few studies directly focused on the relationship between these two markers and SpA. In terms of the inflammatory process, ZFP91 plays a role in the non-canonical NF-kB pathway (35) and is required to maintain regulatory T cell homeostasis (36). With respect to the pathological bone formation, the PUM1 gene was differentially expressed in osteoporosis-related cells (37). Therefore, we speculated that PUM1 and ZFP91 might participate in the process of osteoarticular involvement by activated inflammation or pathological bone formation. Further, we found a significant positive correlation between PUM1 and T-cells CD4 memory resting in psoriatic samples. T-cells CD4 memory resting had a negative correlation with osteoarticular involvement in Ps. We speculate that PUM1 may mediate bone and joint involvement in Ps by inhibiting T-cells CD4 memory resting.

Our study had few limitations. Transcriptome analysis of peripheral blood is a useful approach to compare genotype-similar but phenotype-distinct diseases, or even diseases with different-organ involvement. However, the expression profiling of peripheral blood mononuclear cells should be confirmed by the expression profiling of the target organ. Considering the phenotypic disturbances of the PsA group, the current study is the preliminary exploration of genetic factors related to osteoarthritis involvement in Ps, hoping to provide some meaningful directions for follow-up research.

To conclude, this study is the first to explore the pathways and biomarkers of osteoarticular involvement in psoriasis and AS using the bioinformatics tool. Besides, our study revealed the mRNA surveillance pathway and two diagnostic gene biomarkers

(PUM1 and ZFP91) for the osteoarticular involvement in psoriasis and AS. In addition, by exploration of the two typical diseases AS and PsA, this study may also provide a new perspective to the pathogenesis of SpA.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics committee of Zhujiang Hospital of Southern Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Y-PZ and XW designed and conducted the whole research. YQ and L-GJ, X-TZ collected the GEO datasets and carried out initial data analysis. Y-PZ, XW, JW, and Q-HY completed the data analysis and drafted the manuscript. JW and Q-HY revised and finalized the manuscript. All authors contributed to the article and approved the submitted version. Y-PZ and XW have contributed equally to this work.

FUNDING

This study received the funding of the Guangdong Basic and Applied Basic Research Fund (grant number 2020A1515010039) and the National Natural Science Foundation of China (grant no. 81771727). The National Natural Science Foundation of China (grant no. 82074160).

ACKNOWLEDGMENTS

We thank the authors of the GSE61281, GSE25101, GSE73754, and GSE13355 datasets for their contribution.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.836533/full#supplementary-material

REFERENCES

1. Napolitano M, Caso F, Scarpa R, Megna M, Patri A, Balato N, et al. Psoriatic Arthritis and Psoriasis: Differential Diagnosis. *Clin Rheumatol* (2016) 35:1893–901. doi: 10.1007/s10067-016-3295-9
2. Gottlieb AB, Mease PJ, Mark Jackson J, Eisen D, Amy Xia H, Asare C, et al. Clinical Characteristics of Psoriatic Arthritis and Psoriasis in Dermatologists' Offices. *J Dermatol Treat* (2006) 17(5):279–87. doi: 10.1080/09546630600823369
3. He J, Tang J, Feng Q, Li T, Wu K, Yang K, et al. Weighted Gene Co-Expression Network Analysis Identifies RHOH and TRAF1 as Key Candidate Genes for Psoriatic Arthritis. *Clin Rheumatol* (2021) 40(4):1381–91. doi: 10.1007/s10067-020-05395-8
4. Patrick MT, Stuart PE, Raja K, Gudjonsson JE, Tejasvi T, Yang J, et al. Genetic Signature to Provide Robust Risk Assessment of Psoriatic Arthritis Development in Psoriasis Patients. *Nat Commun* (2018) 9(1):4178. doi: 10.1038/s41467-018-06672-6
5. Stuart PE, Nair RP, Tsoi LC, Tejasvi T, Das S, Kang HM, et al. Genome-Wide Association Analysis of Psoriatic Arthritis and Cutaneous Psoriasis Reveals Differences in Their Genetic Architecture. *Am J Hum Genet* (2015) 97 (6):816–36. doi: 10.1016/j.ajhg.2015.10.019
6. Veale DJ, Fearon U. The Pathogenesis of Psoriatic Arthritis. *Lancet* (2018) 391(10136):2273–84. doi: 10.1016/S0140-6736(18)30830-4
7. Zochling J, Brandt J, Braun J. The Current Concept of Spondyloarthritis With Special Emphasis on Undifferentiated Spondyloarthritis. *Rheumatol (Oxford)* (2005) 44(12):1483–91. doi: 10.1093/rheumatology/kei047
8. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: Archive for Functional Genomics Data Sets–Update. *Nucleic Acids Res* (2013) 41(D1):D991–5. doi: 10.1093/nar/gks1193
9. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res* (2015) 43(7):e47. doi: 10.1093/nar/gkv007
10. Langfelder P, Horvath S. WGCNA: An R Package for Weighted Correlation Network Analysis. *BMC Bioinf* (2008) 9:559. doi: 10.1186/1471-2105-9-559
11. Lin X, Yang F, Zhou L, Yin P, Kong H, Xing W, et al. A Support Vector Machine-Recursive Feature Elimination Feature Selection Method Based on Artificial Contrast Variables and Mutual Information. *J Chromatogr B Analyt Technol BioMed Life Sci* (2012) 910:149–55. doi: 10.1016/j.jchromb.2012.05.020
12. Yoon S, Kim S. AdaBoost-Based Multiple SVM-RFE for Classification of Mammograms in DDSM. *BMC Med Inf Decis Making* (2009) 9:S1. doi: 10.1186/1472-6947-9-S1-S1
13. Walter W, Sánchez-Cabo F, Ricote M. GOplot: An R Package for Visually Combining Expression Data With Functional Analysis. *Bioinformatics* (2015) 31(17):2912–4. doi: 10.1093/bioinformatics/btv300
14. Yu G, Wang LG, Han Y, He QY. Clusterprofiler: An R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J Integr Biol* (2012) 16 (5):284–7. doi: 10.1089/omi.2011.0118
15. Hänzelmann S, Castelo R, Guinney J. GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC Bioinf* (2013) 14:7. doi: 10.1186/1471-2105-14-7
16. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) Hallmark Gene Set Collection. *Cell Syst* (2015) 1(6):417–25. doi: 10.1016/j.cels.2015.12.004
17. Han H, Cho JW, Lee S, Yun A, Kim H, Bae D, et al. TRRUST V2: An Expanded Reference Database of Human and Mouse Transcriptional Regulatory Interactions. *Nucleic Acids Res* (2018) 46(D1):D380–6. doi: 10.1093/nar/gkx1013
18. Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J. NetworkAnalyst 3.0: A Visual Analytics Platform for Comprehensive Gene Expression Profiling and Meta-Analysis. *Nucleic Acids Res* (2019) 47(W1):W234–41. doi: 10.1093/nar/gkz240
19. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling Tumor Infiltrating Immune Cells With CIBERSORT. *Methods Mol Biol* (2018) 1711:243–59. doi: 10.1007/978-1-4939-7493-1_12
20. Sharip A, Kunz J. Understanding the Pathogenesis of Spondyloarthritis. *Biomolecules* (2020) 10(10):1461. doi: 10.3390/biom10101461

21. Rudwaleit M, van der Heijde D, Landewé R, Akkoc N, Brandt J, Chou CT, et al. The Assessment of SpondyloArthritis International Society Classification Criteria for Peripheral Spondyloarthritis and for Spondyloarthritis in General. *Ann Rheum Dis* (2011) 70(1):25–31. doi: 10.1136/ard.2010.133645
22. Ritchlin CT, Colbert RA, Gladman DD. Psoriatic Arthritis. *N Engl J Med* (2017) 376(10):957–70. doi: 10.1056/NEJMra1505557
23. Yao M, Zhang C, Gao C, Wang Q, Dai M, Yue R, et al. Exploration of the Shared Gene Signatures and Molecular Mechanisms Between Systemic Lupus Erythematosus and Pulmonary Arterial Hypertension: Evidence From Transcriptome Data. *Front Immunol* (2021) 12:658341. doi: 10.3389/fimmu.2021.658341
24. Zou R, Zhang D, Lv L, Shi W, Song Z, Yi B, et al. Bioinformatic Gene Analysis for Potential Biomarkers and Therapeutic Targets of Atrial Fibrillation-Related Stroke. *J Transl Med* (2019) 17(1):45. doi: 10.1186/s12967-019-1790-x
25. Sezin T, Vorobyev A, Sadik CD, Zillikens D, Gupta Y, Ludwig RJ. Gene Expression Analysis Reveals Novel Shared Gene Signatures and Candidate Molecular Mechanisms Between Pemphigus and Systemic Lupus Erythematosus in CD4(+) T Cells. *Front Immunol* (2017) 8:1992. doi: 10.3389/fimmu.2017.01992
26. Tew SR, Vasieva O, Peffers MJ, Clegg PD. Post-Transcriptional Gene Regulation Following Exposure of Osteoarthritic Human Articular Chondrocytes to Hyperosmotic Conditions. *Osteoarthr Cartil* (2011) 19 (8):1036–46. doi: 10.1016/j.joca.2011.04.015
27. Valin A, Del Rey MJ, Municio C, Usategui A, Romero M, Fernández-Felipe J, et al. IL6/sIL6R Regulates Tnfα-Inflammatory Response in Synovial Fibroblasts Through Modulation of Transcriptional and Post-Transcriptional Mechanisms. *BMC Mol Cell Biol* (2020) 21(1):74. doi: 10.1186/s12860-020-00317-7
28. Lam IKY, Chow JX, Lau CS, Chan VSF. MicroRNA-Mediated Immune Regulation in Rheumatic Diseases. *Cancer Lett* (2018) 431:201–12. doi: 10.1016/j.canlet.2018.05.044
29. Iwamoto N, Kawakami A. Recent Findings Regarding the Effects of microRNAs on Fibroblast-Like Synovial Cells in Rheumatoid Arthritis. *Immunol Med* (2019) 42(4):156–61. doi: 10.1080/25785826.2019.1695490
30. Rigby RE, Rehwinkel J. RNA Degradation in Antiviral Immunity and Autoimmunity. *Trends Immunol* (2015) 36(3):179–88. doi: 10.1016/j.it.2015.02.001
31. Saalfeld W, Mixon AM, Zelie J, Lydon EJ. Differentiating Psoriatic Arthritis From Osteoarthritis and Rheumatoid Arthritis: A Narrative Review and Guide for Advanced Practice Providers. *Rheumatol Ther* (2021) 8(4):1493–517. doi: 10.1007/s40744-021-00365-1
32. Caputo V, Strafella C, Termine A, Dattola A, Mazzilli S, Lanna C, et al. Overview of the Molecular Determinants Contributing to the Expression of Psoriasis and Psoriatic Arthritis Phenotypes. *J Cell Mol Med* (2020) 24 (23):13554–63. doi: 10.1111/jcmm.15742
33. Leijten EF, van Kempen TS, Olde Nordkamp MA, Pouw JN, Kleinrensink NJ, Vincken NL, et al. Tissue-Resident Memory CD8+ T Cells From Skin Differentiate Psoriatic Arthritis From Psoriasis. *Arthritis Rheumatol* (2021) 73(7):1220–32. doi: 10.1002/art.41652
34. Ezeonyeji A, Baldwin H, Vukmanovic-Stejic M, Ehrenstein MR. CD4 T-Cell Dysregulation in Psoriatic Arthritis Reveals a Regulatory Role for IL-22. *Front Immunol* (2017) 8:1403. doi: 10.3389/fimmu.2017.01403
35. Jin HR, Jin X, Lee JJ. Zinc-Finger Protein 91 Plays a Key Role in LIGHT-Induced Activation of Non-Canonical NF-κb Pathway. *Biochem Biophys Res Commun* (2010) 400(4):581–6. doi: 10.1016/j.bbrc.2010.08.107
36. Wang A, Ding L, Wu Z, Ding R, Teng XL, Wang F, et al. ZFP91 Is Required for the Maintenance of Regulatory T Cell Homeostasis and Function. *J Exp Med* (2021) 218(2):e20201217. doi: 10.1084/jem.20201217
37. Hu Y, Tan LJ, Chen XD, Liu Z, Min SS, Zeng Q, et al. Identification of Novel Potentially Pleiotropic Variants Associated With Osteoporosis and Obesity Using the cFDR Method. *J Clin Endocrinol Metab* (2018) 103(1):125–38. doi: 10.1210/jc.2017-01531

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 04 April 2022
doi: 10.3389/fgene.2022.818994

The Prognostic Significance of RIMKLB and Related Immune Infiltrates in Colorectal Cancers

Yinghao Cao[1†], *Shenghe Deng*[1†], *Lizhao Yan*[2†], *Junnan Gu*[1], *Fuwei Mao*[1], *Yifan Xue*[1], *Le Qin*[1], *Zhengxing Jiang*[1], *Wentai Cai*[3], *Changmin Zheng*[4], *Xiu Nie*[5], *Hongli Liu*[6], *Zhuolun Sun*[7], *Fumei Shang*[8], *Kaixiong Tao*[1], *Jiliang Wang*[1], *Ke Wu*[1]*, *Bin Zhu*[9]* and *Kailin Cai*[1]*

[1]Department of Gastrointestinal Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [2]Department of Hand Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [3]College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, China, [4]School of Optical and Electronic Information, Huazhong University of Science and Technology, Wuhan, China, [5]Department of Pathology, Union Hospital, Tongji Medical, Huazhong University of Science and Technology, Wuhan, China, [6]Cancer Center, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [7]Department of Urology, Third Affiliated Hospital of Sun Yat-sen University, Guangzhou0, China, [8]Department of Medical Oncology, Nanyang Central Hospital, Nanyang, China, [9]Department of Infectious Diseases, Union Hospital, Tongji Medcial College, Huazhong University of Science and Technology, Wuhan, China

OPEN ACCESS

Edited by:
Lam Cheung Tsoi,
University of Michigan, United States

Reviewed by:
Zheng Liu,
National Cancer Center of China, China
Suprabhat Mukherjee,
Kazi Nazrul University, India
Ali Salehzadeh,
Islamic Azad University, Rasht Branch, Iran

***Correspondence:**
Ke Wu
wuke201288@126.com
Bin Zhu
zhubin@hust.edu.cn
Kailin Cai
caikailin@hust.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *20 November 2021*
Accepted: *14 March 2022*
Published: *04 April 2022*

Citation:
Cao Y, Deng S, Yan L, Gu J, Mao F, Xue Y, Qin L, Jiang Z, Cai W, Zheng C, Nie X, Liu H, Sun Z, Shang F, Tao K, Wang J, Wu K, Zhu B and Cai K (2022) The Prognostic Significance of RIMKLB and Related Immune Infiltrates in Colorectal Cancers. Front. Genet. 13:818994. doi: 10.3389/fgene.2022.818994

RimK-like family member B (RIMKLB) is an enzyme that post-translationally modulates ribosomal protein S6, which can affect the development of immune cells. Some studies have suggested its role in tumor progression. However, the relationships among RIMKLB expression, survival outcomes, and tumor-infiltrating immune cells (TIICs) in colorectal cancer (CRC) are still unknown. Therefore, we analyzed RIMKLB expression levels in CRC and normal tissues and investigated the correlations between RIMKLB and TIICs as well as the impact of RIMKLB expression on clinical prognosis in CRC using multiple databases, including the Tumor Immune Estimation Resource (TIMER), Gene Expression Profiling Interactive Analysis (GEPIA), PrognoScan, and UALCAN databases. Enrichment analysis was conducted with the cluster Profiler package in R software to explore the RIMKLB-related biological processes involved in CRC. The RIMKLB expression was significantly decreased in CRC compared to normal tissues, and correlated with histology, stage, lymphatic metastasis, and tumor status ($p < 0.05$). Patients with CRC with high expression of RIMKLB showed poorer overall survival (OS) (HR = 2.5,p = 0.00,042), and inferior disease-free survival (DFS) (HR = 1.9,p = 0.19) than those with low expression of RIMKLB. TIMER analysis indicated that RIMKLB transcription was closely related with several TIICs, including CD4$^+$ and CD8$^+$ T cells, B cells, tumor-associated macrophages (TAMs), monocytes, neutrophils, natural killer cells, dendritic cells, and subsets of T cells. Moreover, the expression of RIMKLB showed significant positive correlations with infiltrating levels of PD1 (r = 0.223, p = 1.31e-06; r = 0.249, p = 1.25e-03), PDL1 (r = 0.223, p = 6.03e-07; r = 0.41, p = 5.45e-08), and CTLA4 (r = 0.325, p = 9.68e-13; r = 0.41, p = 5.45e-08) in colon and rectum cancer, respectively. Enrichment analysis showed that the RIMKLB expression was positively related to extracellular matrix and immune inflammation-related pathways. In conclusion, RIMKLB expression is associated with survival outcomes and TIICs levels in patients with CRC, and therefore, might be a potential novel prognostic biomarker that reflects the immune infiltration status.

Keywords: colorectal cancer, RIMKLB, tumor-infiltrating immune cells, prognosis, biomarker

INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer-related death (Siegel et al., 2021). The global incidence of CRC is expected to increase to 2.5 million new cases by 2035, with a steady and declining trend only in highly developed countries (Bray et al., 2018). Furthermore, studies show that CRC is now beginning to develop at a younger age (Dekker et al., 2019; Mauri et al., 2019). In CRC management, metastasis is an important biological feature leading to poor prognosis. Immune-related mechanisms play an important role in digestive tract cancer, especially in CRC (Ganesh et al., 2019; Siegel et al., 2019). In the past decade, a high tumor mutation burden has become a hallmark of immunotherapeutic response in some tumor types due to the success of immunotherapy in achieving long-lasting responses in previously difficult-to-treat solid tumors (Samstein et al., 2019). However, the clinical efficacy of CRC immunotherapy in metastatic CRC is poor, and the popularly used anti-PD-1 and anti-PD-L1 show partial reactions in metastatic CRC and gastric cancer (Galon et al., 2007; Rahma and Hodi, 2019). In these tumors, low tumor mutation load and lack of immune cell infiltration are thought to be mechanisms of immune resistance (Galon et al., 2007). In addition, increasingly more and more studies have found that tumor immune cell infiltration is closely related to the prognosis and efficacy of CRC chemotherapy and immunotherapy (Rahma and Hodi, 2019). Therefore, it is of great significance to elucidate the immunophenotype of CRC-immune interaction, the mechanism of immunotherapeutic resistance, and the identification of new immune-related therapeutic targets.

RIMK is a unique protein that in *Escherichia* coli that acts as an ATP-dependent enzyme that induces oligo-glutamylation of ribosomal protein S6 (S6) after transcription, and bacterial S6 is the target of oligo-glutamylation of ATP-dependent glutamate ligase RIMK (Kino et al., 2011; Pletnev et al., 2019). In *Pseudomonas aeruginosa*, the lack of RIMK can shorten its survival time owing to the functional effect of RIMK on ribosome properties (Grenga et al., 2017). RIMKLB is a mammalian homologous gene of RIMK, it has been cloned in mammals, resulting in β-citrylglutamate (β-CG) or N-acetylaspartylglutamate synthase activity (Collard et al., 2010). Some studies have found that RIMKLB can affect reproductive function of mammals, and in tumor research, RIMKLB may coordinate with DDIT4 function to mediate mTOR inhibition and growth inhibition of tumor cells (Wang et al., 2015; Maekura et al., 2021). Some studies have found that RIMKL modeling can accurately predict the 5-years survival rate of patient with colon cancer patients, suggesting that RIMKL may play a role in tumor progression (Huang et al., 2021). However, the specific role of RIMKLB in CRC is unknown and needs further study.

The tumor microenvironment and tumor-infiltrating immune cells play an important role in CRC tumor progression. The immune components of the tumor microenvironment can regulate tumor progression and are attractive therapeutic targets. A large number of studies have shown that high a infiltration rate of $CD8^+$ and $CD4^+$ T cells is associated with better prognosis in patients with CRC patients (Tada et al., 2016). Furthermore, high infiltration of dendritic cells (DCs) in tumors has been reported to be associated with more favorable clinical outcomes (Gulubova et al., 2012). Some studies have also shown that extensive infiltration of NK cells in tumors has a good prognostic effect on CRC (Bindea et al., 2013). However, there are no studies have been reported on RIMKLB and the immune microenvironment in CRC, and it remains unknown whether RIMKLB can affect immune cells and tumor microenvironment and promote tumor progression.

Therefore, based on a list of public databases, our study aimed to determine the correlation between RIMKLB expression and tumor-infiltrating immune cells (TIICs) in CRC. Moreover, we also performed subgroup analysis *via* tumor site to determine whether the role of RIMKLB in colon cancer is different from that in rectum cancer.

MATERIALS AND METHODS

UALCAN and Tumor Immune Estimation Resource Database Analysis

The expression level of the RIMKLB gene in various types of cancers was identified in the UALCAN (http://ualcan.path.uab.edu/cgi-bin/ualcan-res-prot.pl) (Chandrashekar et al., 2017) and TIMER database (https://cistrome.shinyapps.io/timer/) (Li et al., 2017). In addition, we focused on an easy-to-use webtool, GEPIA, which is available at http://gepia.cancer-pku.cn/index.html, to study the differential expression of RIMKLB mRNA in CRC tissues and normal tissues (Tang et al., 2017).

GEPIA and PrognoScan Database Analysis

Using logarithmic rank test, GEPIA was used to generate survival curves, including overall survival (OS) and disease-free survival (DFS), based on gene expression in colon and rectal cancer. The association between RIMKLB expression and OS in CRC was analyzed *via* PrognoScan database (http://www.abren.net/PrognoScan/) (Mizuno et al., 2009), whose data are different from that of The Cancer Genome Atlas (TCGA) database. The threshold was adjusted to a Cox p-value < 0.05.

Tumor Immune Estimation Resource Database

The TIMER database includes 10,897 samples across 32 cancer types based on RNA-Seq expression profiling data from TCGA database. It can test the differential gene expression in tumor tissues, the abundance of TIICs from gene expression profiles, and the statistical correlation between the two genes by the statistical method through gene expression data (Li et al., 2016). Therefore, we analyzed the relationship between RIMKLB expression and TIICs, including $CD4^+$ and $CD8^+$ T cells, B cells, neutrophils, DCs, and macrophages.

Additionally, the correlation between RIMKLB expression and gene markers of TIICs, including $CD8^+$ T cells, T cells (general), B cells, monocytes, tumor-associated macrophages (TAMs), M1 macrophages, M2 macrophages, neutrophils, natural killer (NK) cells, DCs, T-helper 1 (Th1) cells, T-helper 2 (Th2) cells, follicular helper T (Tfh) cells, T-helper 17 (Th17) cells, Tregs, and exhausted T cells, was explored through related modules, which were reported in a previous study (Wu et al., 2020).

Gene Correlation Identification in GEPIA

The GEPIA database contains the gene expression data from 8,587 normal and 9,736 tumor tissue samples of TCGA and the Genotype-Tissue Expression (GTEx) projects, and can be used to further identify the significantly correlated genes in TIMER (Tang et al., 2017). GEPIA was also used to generate survival curves and determine OS and DFS rates, differential gene expression, and the relationship between two genes. The spearman method was used to determine the correlation coefficient, and a median value of the RIMKLB expression was used as a cutoff to distinguish high expression from low expression.

Oncogenomics and Mutational Study

We use cBioPortal6 to analyze the impact of the RIMKLB gene in the Colorectal Adenocarcinoma TCGA PanCancer dataset containing 594 samples. Further using the mRNA expression data of the top 25 positively correlated genes to indicate the correlated gene with RIMKLB in CRC. The cancer type summary tab provides a detailed overview of the RIMKLB gene in different subtypes of CRC, i.e., mucinous adenocarcinoma of colon and rectum, colon adenocarcinoma, and rectal adenocarcinoma. It also showed mutations in CRC's RIMKLB gene and mutations within the associated genome. Different types of mutations associated with the RIMKLB gene in CRC were analyzed using COSMIC-"Catalogue of Somatic Mutations in Cancer."

Enrichment Analysis

Patients with CRC were initially divided into high RIMKLB expression and low expression groups. Genes that were differentially expressed between the two groups were screenedto explore the functional role of RIMKLB in CRC with the false discovery rate less than 0.05, and $|\log FC| \geq 1$ combined with p value less than 0.05 were regarded as significant.

Statistical Analysis

The data were analyzed using the GraphPad Prism (version 6.0) and SPSS (version 21.0). Low and high RIMKLB groups were established based on the median expression of RIMKLB transcription in the separate datasets. Survival curves were generated from the PrognoScan, Kaplan-Meier plots and GEPIA database. The relation of RIMKLB expression and TICSs was evaluated by Spearman's correlation, and the strength of the correlation was determined using the following guide for the absolute value: 0.00–0.29 (weak), 0.30–0.59 (moderate), 0.60–0.79 (strong), 0.80–1.00 (very strong) (Gao et al., 2017). p-values <0.05 were considered statistically significant.

RESULT

The Expression Levels Analysis of RimK-Like Family Member B in Different Types of Human Cancers and Normal Tissues

To determine the difference expression of RIMKLB between tumor and normal tissue, we used the UALCAN database to analyze the expression levels of RIMKLB in normal tissues of different tumors and multiple cancer types. The result showed that RIMKLB expression was lower in bladder, breast, cervical, cloln, rectum, glioblastoma, lung, pancreatic, prostate, thyroid, thymoma, stomach, uterine corpus endometrial carcinoma, and it was higher in cholangio, esophageal, head and neck, kidney, pheochromocytoma and paraganglioma, sarcomav (**Figure 1A**). TIMER database was utilized to validate the expression profiles of RIMKLB in pancancer, and RIMKLB mRNA was also lowly expressed in CRC tissues (**Figure 1B**). The GEPIA database was used to analyze the expression of RIMKLB TPM in colon cancer (**Figure 1C**) and rectum cancer (**Figure 1D**). Red represents colon cancer tissue; purple represents normal colon tissue, which is statistically significant ($p < 0.05$).

Relationship Between RimK-Like Family Member B Expression and Clinicopathological Characteristics of Patients With Colorectal Cancer

To investigate the relationship between mRNA expression of RIMKLB and clinicopathological features of CRC patients, we analyzed clinical information from CRC samples from the TCGA project. The results (**Figure 2**) revealed that the mRNA expression of RIMKLB was significantly increased in the mucinous adenocarcinoma ($p < 0.001$), rectum ($p < 0.001$), lymph node stage (N0) ($p = 0.0485$), advanced stages (III/IV) ($p < 0.001$), and with tumor ($p < 0.001$). However, there was no significant correlation between RIMKLB mRNA expression and gender ($p = 0.5223$), age ($p = 0.6097$), advanced tumor ($p = 0.909$) and metastasis status ($p = 0.921$).

Prognostic Significance of RimK-Like Family Member B Expression in Colorectal CanceC

The prognostic significance of RIMKLB expression in CRC was analyzed using the TCGA RNA sequencing data from the GEPIA database. High RIMKLB expression levels were associated with poorer

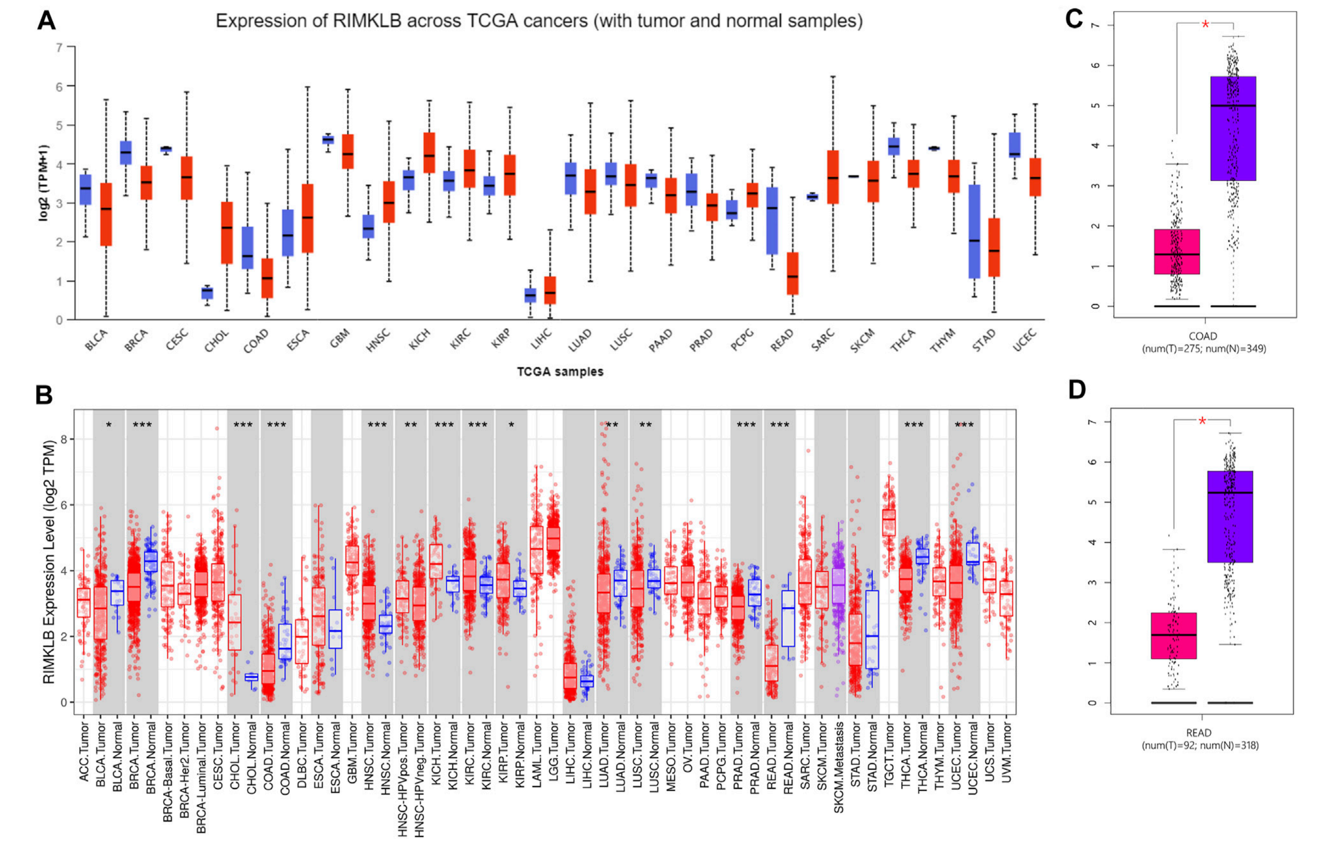


FIGURE 1 | RIMKLB expression levels in different types of human cancers. **(A)** Increased or decreased RIMKLB in data sets of different cancers compared with normal tissues in the UALCAN database; **(B)** Human RIMKLB expression levels in different tumor types from TCGA database were determined by TIMER (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$); **(C)** Decreased RIMKLB in colon cancer compared with normal tissues; **(D)** Decreased RIMKLB in rectum cancer compared with normal tissues.

OS (HR = 2.3, p = 0.0003, **Figure 3A**), and DFS in CRC (HR = 2, p = 0.0012, **Figure 3B**). When subgrouped by tumor site, this association only existed in colon cancer (OS: HR = 2.5, P = 0.00042, **Figure 3C**; DFS: HR = 2.5, P = 0.00028, **Figure 3D**) and disappeared in rectal cancer (OS: HR = 1.5, P = 0.39, **Figure 3E**; DFS: HR = 1.9, P = 0.19, **Figure 3F**).

We also verified the prognostic value of RIMKLB expression in CRC cancers using the Prognoscan website, whose data were from GEO database. High RIMKLB expression was associated with worse OS (HR = 2.63, 95% CI = 1.38–5.02, p = 0.0034, **Figure 4A**) among CRC patients in GSE17536, this survival significance (HR = 5.6, 95% CI = 1.24–25.36, p = 0.0255, **Figure 4B**) was also observed in GSE17537. In brief, high expression of RIMKLB is a potent risk factor among CRC patients.

RimK-Like Family Member B Expression Levels Correlate With the Infiltration Levels of Immune Cells in Colorectal Cance

Previous studies have reported that survival time for colorectal cancers depends on the number and activity of tumor-infiltrating lymphocytes (Ohtani, 2007; Japanese Gastric Cancer Association, 2017). Therefore, we explored the relationship between RIMKLB expression with prognosis and the infiltrating immune cells in CRC using the TIMER and GEPIA database.

The level of IMKLB expression positively correlated with the infiltration levels of $CD8^+$ T cells (r = 0.131, p = 8.04e-03), $CD4^+$ T (r = 0.428, p = 2.27e-19) cells, macrophages (r = 0.463, p = 7.54e-23), neutrophils (r = 0.315, p = 1.03e-10), and dendritic cells (r = 0.355, p = 2.03e-13), but negatively related to tumor purity (r = –0.266, p = 5.16e-08) and B cells (r = –0.044, p = 3.57e-01) in Colon adenocarcinoma (COAD) tissues (**Figure 5A**); The level of RIMKLB expression is significantly negatively related to tumor purity (r = –0.298, p = 3.47e-04) and has significant positive correlations with infiltrating levels of B cells (r = 0.151, p = 7.56e-02), $CD8^+$ T cells (r = 0.229, p = 6.68e-03), $CD4^+$ T cells (r = 0.347, p = 2.90e-05), macrophages (r = 0.307, p = 2.37e-04), neutrophils (r = 0.227, p = 7.31e-03), and dendritic cells (r = 0.322, p = 1.11e-04) in Rectum adenocarcinoma (READ) tissues (**Figure 5B**).

Correlation of RimK-Like Family Member B Expression With Immune Checkpoint in COAD and READ.

Research on the role of targeted immune therapy in the treatment of advanced colorectal cancer and its relationship

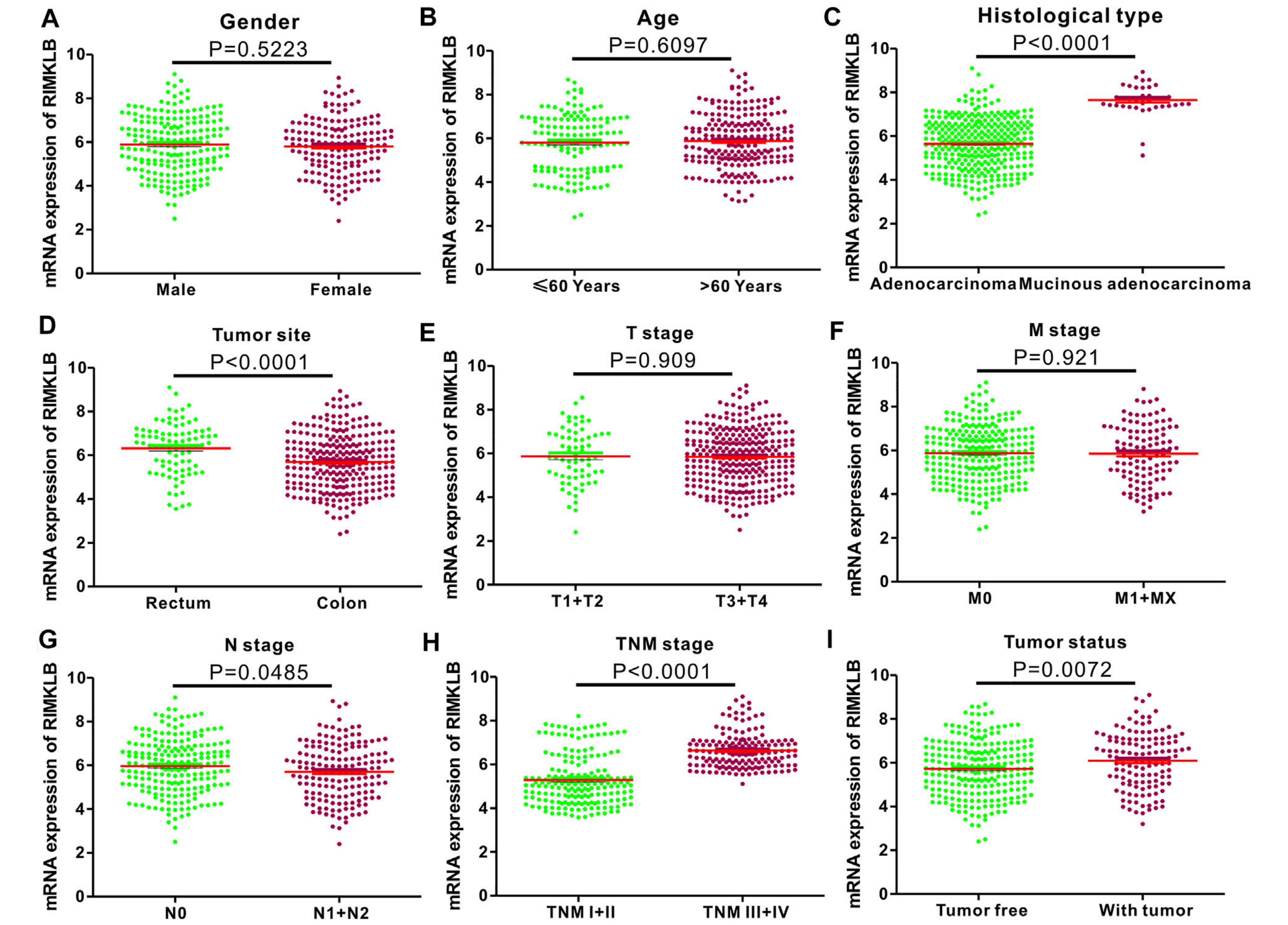


FIGURE 2 | Correlation between RIMKLB mRNA expression and clinical indexes of CRC patients from TCGA database. RIMKLB mRNA expression is stratified with scatter plots using the TCGA dataset by gender **(A)**, gender **(B)**, histological type **(C)**, Tumor site **(D)**, T stage **(E)**, M stage **(F)**, N stage **(G)**, TNM stage **(H)**, and Tumor status **(I)**.

with tumor gene mutations has received more and more attention (Galon et al., 2007). Therefore, we used TIMER database to investigate the relationship between RIMKLB expression and immunotherapeutic targets in colorectal cancer. The results showed that the expression of RIMKLB was significantly correlated with it. We find that RIMKLB expression has significant positive correlations with infiltrating levels of PD1 (A, r = 0.223, p = 1.31e-06; B, r = 0.16, p = 1.20e-03), PDL1 (C, r = 0.223, p = 6.03e-07; D, r = 0.187, p = 1.47e-04) and CTLA4 (E, r = 0.325, p = 9.68e-13; F, r = 0.265, p = 6.07e-08) before and after purity adjustment in COAD; In addition, RIMKLB expression also has significant positive correlations with infiltrating levels of PD1 (G, r = 0.249, p = 1.25e-03; H, r = 0.121, p = 0.156e-01), PDL1 (I, r = 0.372, p = 9.52e-07; J, r = 2.94, p = 4.50e-04) and CTLA4 (K, r = 0.41, p = 5.45e-08; L, r = 0.284, p = 7.19e-04) before and after purity adjustment in READ (**Figure 6A–L**).

Correlation Between RimK-Like Family Member B mRNA Levels and Different Subsets of Immune Cells

We used TIMER and GEPIA databases to investigate the relationship between RIMKLB and various immune-infiltrating cells based on the expression levels of immune marker genes in colon and rectum tissues. The immune cells analyzed in CRC tissues included CD8$^+$ T cells, CD4$^+$ T cells, B cells, tumor-associated macrophages (TAMs), monocytes, M1 and M2 macrophages, neutrophils, and natural killer (NK) cells, dendritic cells (DCs), subsets of T cells [T helper 1 (Th1), Th2, follicular helper T (Tfh), Th17]. The results showed that RIMKLB expression level was significantly correlated with most immune marker groups of various immune cells and different T cells in colon and rectum cancer (**Table 1**). Interestingly, we found that the expression levels of marker sets of Neutrophils, Th1, M2 macrophages, TAMs, Dendritic cell, monocytes, and

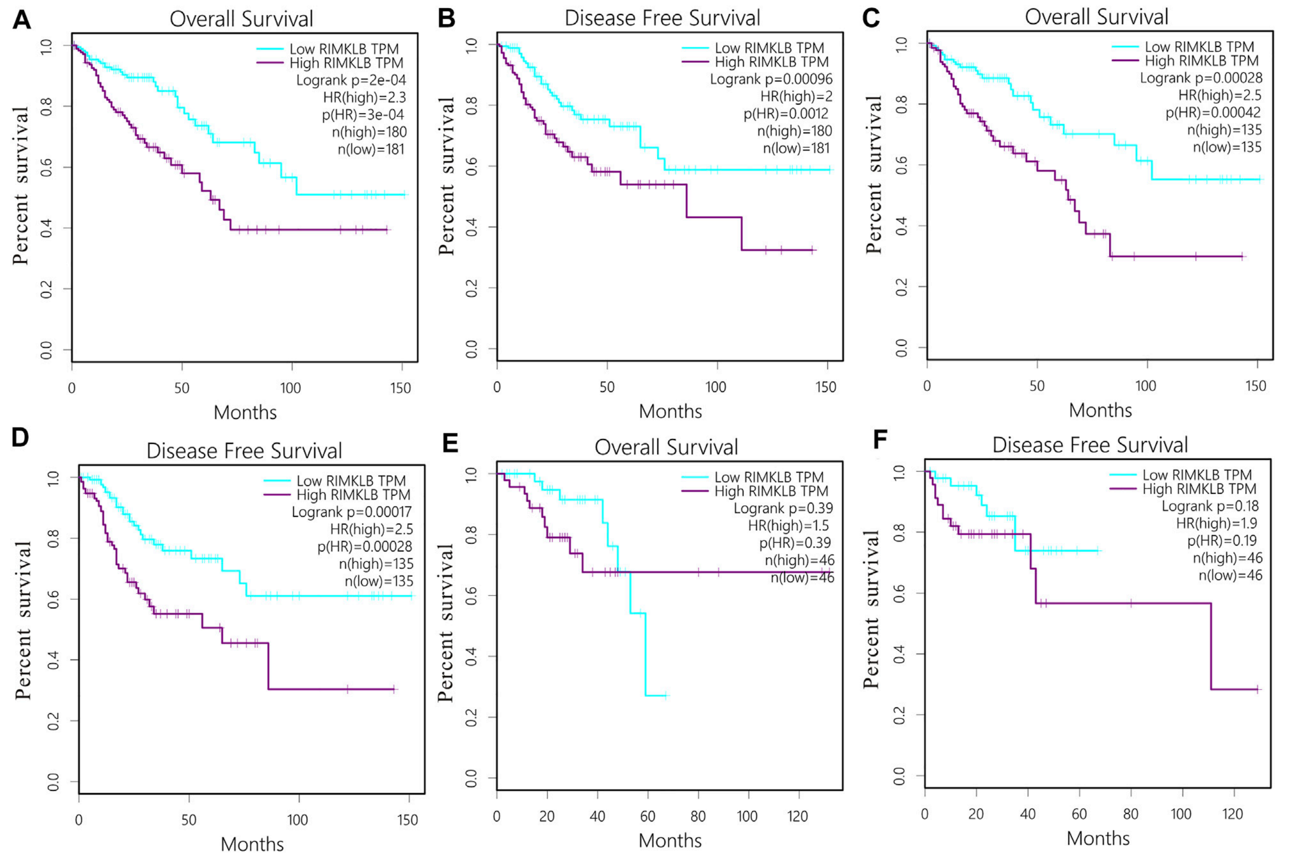


FIGURE 3 | Survival curves and subgroup analysis of patients with colorectal cancer from TCGA cohort. Overall survival (OS) and disease free (DFS) curves comparing the high and low expression of RIMKLB in CRC **(A,B)**; OS and DFS curves comparing the high and low expression of RIMKLB in colon cancer **(C,D)** and rectum cancer **(E,F)**.

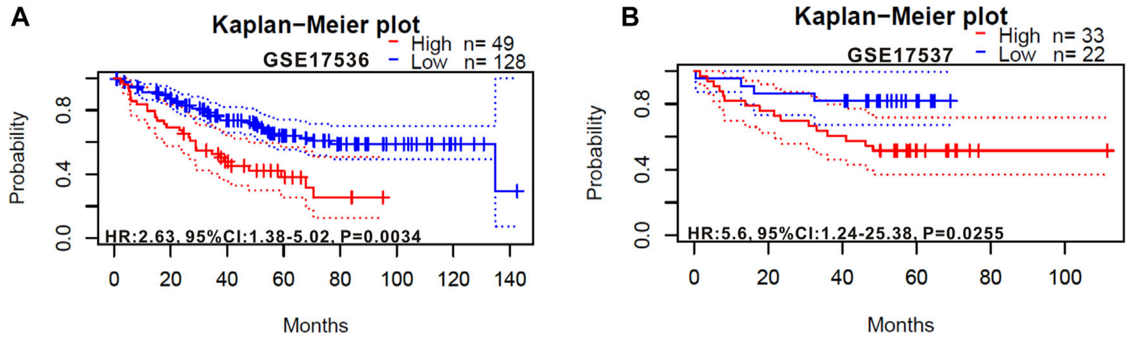


FIGURE 4 | Kaplan-Meier survival curves comparing the high and low expression of RIMKLB in colorectal cancer in the PrognoScan. **(A)** Survival curves of OS colorectal cancer cohorts of GSE17536 (*n* = 177); **(B)** Survival curves of OS colorectal cancer cohorts of GSE17537 (*n* = 55).

Th2 have strong correlations with RIMKLB expression in colon and rectum. The correlation analysis was adjusted for purity because tumor purity of clinical samples affected the analysis of immune infiltration (**Table 2**). To be specific, we showed that ITGAM, and CCR7 of Neutrophils; TBX21, STAT1, STAT4, and TNF of Th1; CD86 and CSF1R of monocytes; CD163, and VSIG4 of M2 macrophages; CCL2, CD68, and IL10 of TAMs (tumor-associated macrophages); HLA-DPB1, HLA-DRA, HLA-DPA1, CD1C, NRP, and ITGAX of Dendritic cell; GATA3, GATA6, and GATA5A of Th2 were significant correlated with RIMKLB expression in COAD ($p < 0.0001$; **Figures 7A–G**) and READ ($p < 0.0001$; **Figures 8A–G**).

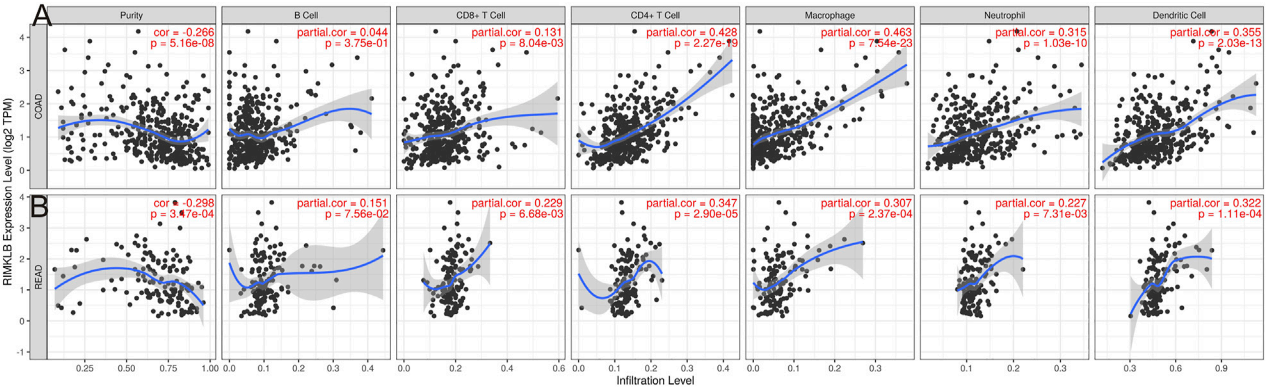


FIGURE 5 | Correlation of RIMKLB expression with immune infiltration level in COAD (colon adenocarcinoma), and READ (recutm adenocarcinoma) in the TIMER database. **(A)** RIMKLB expression is significantly negatively related to tumor purity and has significant positive correlations with infiltrating levels of CD8$^+$ T cells, CD4$^+$ T cells, macrophages, neutrophils, and dendritic cells in COAD, other than B cells. **(B)** RIMKLB expression is significantly negatively related to tumor purity and has significant positive correlations with infiltrating levels of B cells, CD8$^+$ T cells, CD4$^+$ T cells, macrophages, neutrophils, and dendritic cells in READ.

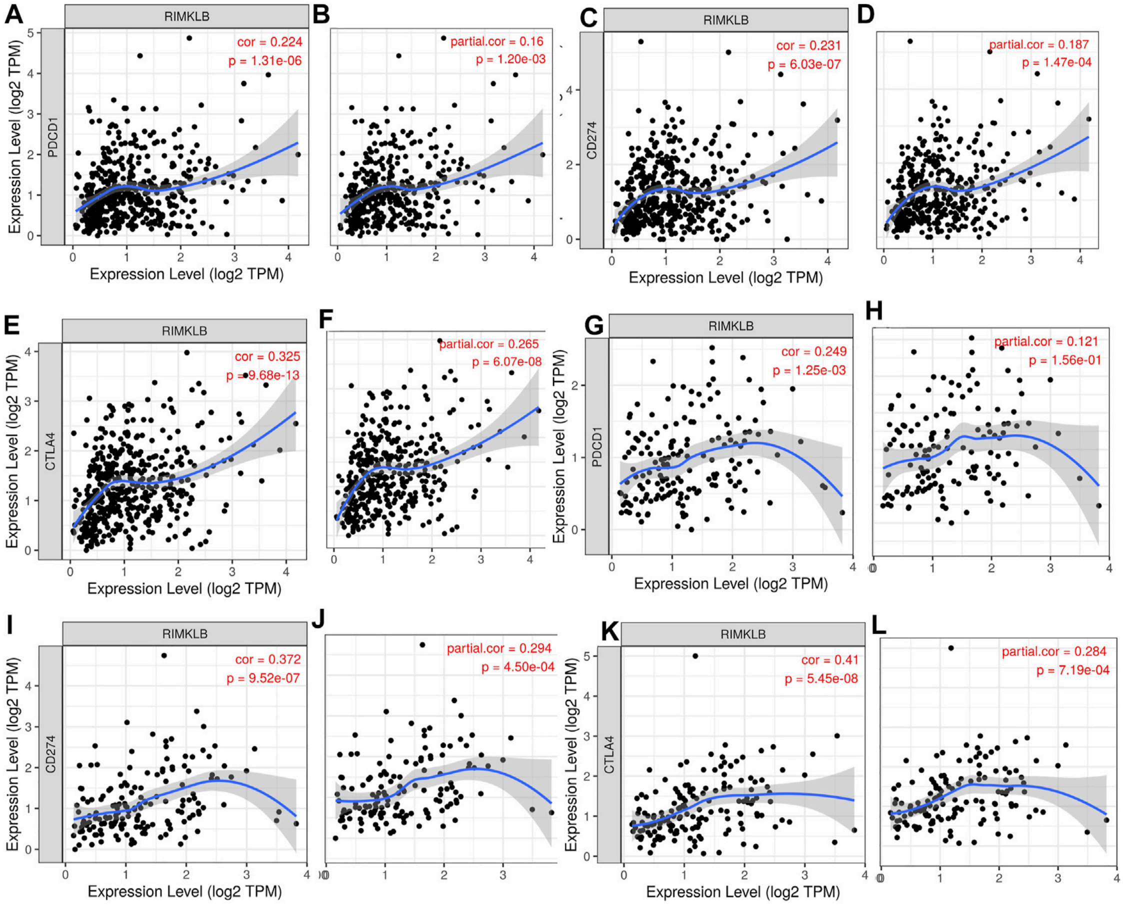


FIGURE 6 | Correlation of RIMKLB expression with PD1, PDL1, and CTLA4 in COAD, and READ. RIMKLB expression has significant positive correlations with infiltrating levels of PD1 **(A,B)**, PDL1 **(C,D)** and CTLA4 **(E,F)** before and after purity adjustment in COAD; As well as in READ [PD1 **(G,H)**, PDL1 **(I,J)** and CTLA4 **(K,L)**].

TABLE 1 | Correlation analysis between RIMKLB and related gene markers of immune cells in COAD and READ.

Cell Type	Marker	COAD				READ			
		None		Purity		None		Purity	
		Cor*	*p* Value	Cor	*p* Value	Cor*	*p* Value	Cor	*p* Value
CD8$^+$ T cell	CD8A	0.203	1.22E-05	0.129	9.43E-03	0.342	7.64E-06	0.237	5.00E-03
	CD8B	0.178	1.34E-04	0.157	1.51E-03	0.204	8.49E-03	0.112	1.90E-01
T cell (general)	CD3D	0.202	1.34E-05	0.131	8.16E-03	0.191	1.39E-02	0.048	5.71E-01
	CD3E	0.28	1.29E-09	0.218	9.24E-06	0.285	2.10E-04	0.159	6.18E-02
	CD2	0.259	2.12E-08	0.185	1.73E-04	0.321	2.85E-05	0.206	1.51E-02
B cell	CD19	0.233	4.42E-07	0.15	2.43E-03	0.087	2.632–01	0.01	9.11E-01
	CD79A	0.278	1.62E-09	0.186	1.59E-04	0.194	1.22E-02	0.054	5.28E-01
Monocyte	CD86	0.404	0.00E + 00	0.357	1.11E-13	0.478	9.65E-11	0.395	1.52E-06
	CD115	0.491	0.00E + 00	0.474	4.09E-24	0.423	1.95E-08	0.332	6.68E-05
TAM	CCL2	0.463	0.00E + 00	0.422	5.31E-19	0.515	0.00E + 00	0.438	6.71E-08
	CD68	0.32	2.16E-12	0.289	3.15E-09	0.357	2.80E-06	0.277	9.52E-04
	IL10	0.324	1.20E-12	0.29	2.47E-09	0.233	2.57E-03	0.151	7.60E-02
M1 macrophage	INOS	−0.194	3.09E-05	−0.23	2.26E-06	−0.085	2.78E-01	−0.044	6.05E-01
	IRF5	0.336	1.92E-13	0.354	1.91E-13	0.237	2.16E-03	0.203	1.66E-02
	COX2	0.187	5.59E-05	0.12	1.54E-02	0.256	9.20E-04	0.159	6.11E-02
M2 macrophage	CD163	0.474	0.00E + 00	0.451	9.60E-22	0.529	0.00E + 00	0.452	2.34E-08
	VSIG4	0.443	0.00E + 00	0.414	2.75E-18	0.314	4.20E-05	0.22	9.12E-03
	MS4A4A	0.403	0.00E + 00	0.365	3.06E-14	0.455	1.11E + 09	0.093	4.33E-01
Neutrophils	CD66b	−0.111	1.73E-02	−0.12	1.67E-02	−0.379	4.84E-07	−0.3	3.47E-04
	CD11b	0.468	0.00E + 00	0.445	3.50E-21	0.526	0.00E + 00	0.434	9.43E-08
	CCR7	0.338	1.34E-13	0.271	2.78E-08	0.187	1.61E-02	0.121	1.56E-01
Natural killer cell	KIR2DL1	0.08	8.72E-02	0.034	4.95E-01	0.186	1.66E-02	0.166	5.02E-02
	KIR2DL3	0.085	6.85E-02	0.075	1.31E-01	0.182	1.88E-02	0.174	4.07E-02
	KIR2DL4	−0.023	6.23E-01	−0.1	4.33E-02	0.1	2.01E-01	−0.032	7.12E-01
	KIR3DL1	0.104	2.60E-02	0.051	3.04E-01	0.08	3.08E-01	0.041	6.34E-01
	KIR3DL2	0.136	3.63E-03	0.076	1.28E-01	0.199	1.01E-02	0.114	1.83E-01
	KIR3DL3	−0.061	1.89E-01	−0.06	2.38E-01	−0.051	5.11E-01	−0.099	2.48E-01
	KIR2DS4	0.033	4.86E-01	0.016	7.47E-01	0.065	4.07E-01	−0.027	7.48E-01
Dendritic cell	HLA-DPB1	0.37	2.33E-16	0.317	6.40E-11	0.331	1.52E-05	0.211	1.27E-02
	HLA-DQB1	0.218	2.61E-06	0.158	1.40E-03	0.097	2.12E-01	0.041	6.35E-01
	HLA-DRA	0.271	4.33E-09	0.21	1.96E-05	0.325	2.11E-05	0.211	1.25E-02
	HLA-DPA1	0.32	2.29E-12	0.263	7.80E-08	0.355	3.07–06	0.228	6.97E-03
	BDCA-1	0.339	9.49E-14	0.296	1.17E-09	0.195	1.16E-02	0.069	4.17E-01
	BDCA-4	0.546	0.00E + 00	0.514	1.05E-28	0.672	0.00E + 00	0.619	4.48E-16
	CD11c	0.484	0.00E + 00	0.454	4.51E-22	0.496	5.97E-12	0.432	1.12E-07
Th1	T-bet	0.267	6.68E-09	0.226	4.43E-06	0.367	1.18E-06	0.281	8.06E-04
	STAT4	0.308	2.00E-11	0.267	4.70E-08	0.332	1.37E-05	0.272	1.21E-03
	STAT1	0.258	2.02E-08	0.224	5.36E-06	0.445	2.84E-09	0.363	1.12E-05
	IFN-γ	0.077	9.99E-02	0.041	4.11E-01	0.301	7.96E-05	0.209	1.34E-02
	TNF-α	0.21	5.90E-06	0.17	5.77E-04	0.226	3.41E-03	0.131	1.23E-01
Th2	GATA3	0.486	1.52E-28	0.439	1.61E-20	0.422	1.99E-08	0.337	4.90E-05
	STAT6	0.247	8.08E-08	0.253	2.36E-07	0.181	2.01E-02	0.206	1.49E-02
	STAT5A	0.271	4.36E-09	0.273	2.17E-08	0.137	7.94E-02	0.122	1.53E-01
	IL13	0.161	5.62E-04	0.104	3.59E-02	0.127	1.02E-01	−0.008	9.22E-01
Tfh	BCL6	0.459	0.00E + 00	0.425	3.27E-19	0.597	0.00E + 00	0.595	1.18E-14
	IL21	0.102	2.85E-02	0.066	1.81E-01	0.061	4.43E-01	0.036	6.73E-01
Th17	STAT3	0.246	1.02E-07	−0.27	5.16E-08	0.403	9.86E-08	0.339	3.47E-04
	IL17A	−0.157	7.46E-04	−0.16	1.35E-03	−0.21	6.60E-03	−0.186	2.82E-02

Cor: Correlation.*

Co-Expression and Correlation Amongst the Other Genes Associated With RimK-Like Family Member B in Colorectal Cancer

The top 25 positively co-expressed genes were analyzed *via* cBioPortal, containing the Spearman's correlation coefficient, *p*-value from two-sided *t*-test, and also *q*-value derived from the Benjamini–Hochberg FDR correction procedure (**Table 3**). The correlation graph was obtained using the Pearson's correlation coefficient amongst RIMKLB gene with AKT3 (*r*-value- 0.68), MPDZ (*r*-value-0.66), PKD2 (*r*-value- 0.67) and MAP1B (R-value- 0.69) (**Supplementary Figures S1A–S1D**). Collectively all these results reveal that the RIMKLB gene has a positive association and correlation with AKT3, MPDZ, PKD2, and MAP1B to upregulate the gene expression to induce the development of colorectal cancer.

TABLE 2 | Correlation analysis between RIMKLB and significant gene markers of immune cells in GEPIA.

Description	Gene Markers	COAD		READ	
		Correlation	*p* Value	Correlation	*p* Value
Monocyte	CD86	0.42	2.20E-13	0.27	0.01
	CD115 (CSF1R)	0.51	0	0.29	0.0046
TAM	CCL2	0.53	0	0.25	0.015
	CD68	0.27	5.30E-06	0.19	0.076
	IL10	0.38	7.40E-11	0.2	0.056
M2 Macrophage	CD163	0.43	1.70E-13	0.3	0.0036
	VSIG4	0.44	3.20E-14	0.23	0.03
	MS4A4A	0.47	0	0.26	0.013
Neutrophils	CD11b (ITGAM)	0.47	4.40E-16	0.25	0.018
	CCR7	0.47	0	0.078	0.46
Dendritic cell	HLA-DPB1	0.31	2.00E-07	0.24	0.023
	HLA-DRA	0.25	1.90E-05	0.19	0.075
	HLA-DPA1	0.29	1.10E-06	0.2	0.057
	BDCA-1(CD1C)	0.41	1.60E-12	0.055	0.6
	BDCA-4(NRP1)	0.58	0	0.3	0.0041
	CD11c (ITGAX)	0.42	6.70E-13	0.21	0.044
Th1	T-bet (TBX21)	0.31	1.40E-07	0.18	0.089
	STAT4	0.45	7.30E-15	0.2	0.06
	STAT1	0.22	0.00027	0.19	0.072
	TNF-α (TNF)	0.23	0.00013	0.14	0.18
Th2	GATA3	0.58	0	0.18	0.085
	STAT6	0.15	0.01	0.26	0.013
	STAT5A	0.35	1.70E-09	0.085	0.42

The Mutational Analysis of RimK-Like Family Member B in Colorectal Cancer

The RIMKLB gene mutation was analyzed on COSMIC database comprising more than 2,406 samples of colorectal cancer out of which 77 were recorded for mutations, among them the missense substitution is highest with 53.25% followed by synonymous substitution (23.38%), nonsense substitution (1.30%) and other types (6.49%) (**Supplementary Figure S2A**).

The breakdown of various substitution mutation is shown in **Supplementary Figure S2B**, representing the highest type of G > A (39.66%) and lowest showing T > A (3.45%).

To determine and analyze the frequency and type of mutation, cBioPortal server was used where the cancer type summary indicates the mutation along with the various subtypes of colorectal cancer showing mucinous adenocarcinoma of colon and rectum (>6%), colon adenocarcinoma (<6%), and rectal adenocarcinoma (~2%) (**Supplementary Figure S2C**). The Oncoprint and Mutation tab shows that the RIMKLB gene is altered in 2.5% of the total patients in TCGA colorectal cancer dataset along with the heatmap for the associated genes (**Supplementary Figure S2D**). Additionally, a mutational study for the correlation among the RIMKLB gene with AKT3, MPDZ, PKD2, and MAP1B (**Supplementary Figures S3A–S3D**) showing a significant coefficient value for both Spearman and Pearson Correlation test and the regression line. It is observed that the mutation of RIMKLB is much more expressive for AKT3 > MPDZ > PKD2 > MAP1B.

RimK-Like Family Member B-Related Biological Pathways in Patients With Colorectal Cancer

We carried out the biological process and KEGG pathway to further investigate the potential pathways of RIMKLB in CRC. RIMKLB was mainly involved in cell-cell adhesion *via* plasma-membrane adhesion molecules, humoral immune response mediated by circulating immunoglobulin, cytolysis, killing by host of symbiont cells, triglyceride-rich lipoprotein particle remodeling, regulation of intestinal absorption, chylomicron assembly (**Supplementary Figure S6A**). Moreover, KEGG analysis revealed that RIMKLB is involved in pathways of ECM-receptor interaction, Cell adhesion molecules, Platelet activation, Chemical carcinogenesis-DNA adducts, cAMP signaling pathway, PI3K-Akt signaling pathway, and Cytokine-cytokine receptor interaction. RIMKLB is associated with local immunity in colorectal cancer, and its abnormal expression may lead to the occurrence and development of CRC (**Supplementary Figure S5B**).

DISSCUSSION

This study was the first to reveal the expression and prognostic efficacy of RIMKLB in CRC. We found that the expression of this gene was significantly different in a variety of tumors. Notably, it was significantly decreased in CRC tumors compared to normal tissues, and this was correlated with histology, stage, lymph node metastasis, and tumor status. Moreover, multiple databases confirmed that a high expression of RIMKLB was associated with worse OS and DFS, indicating that this gene may play an important role in tumor development. The

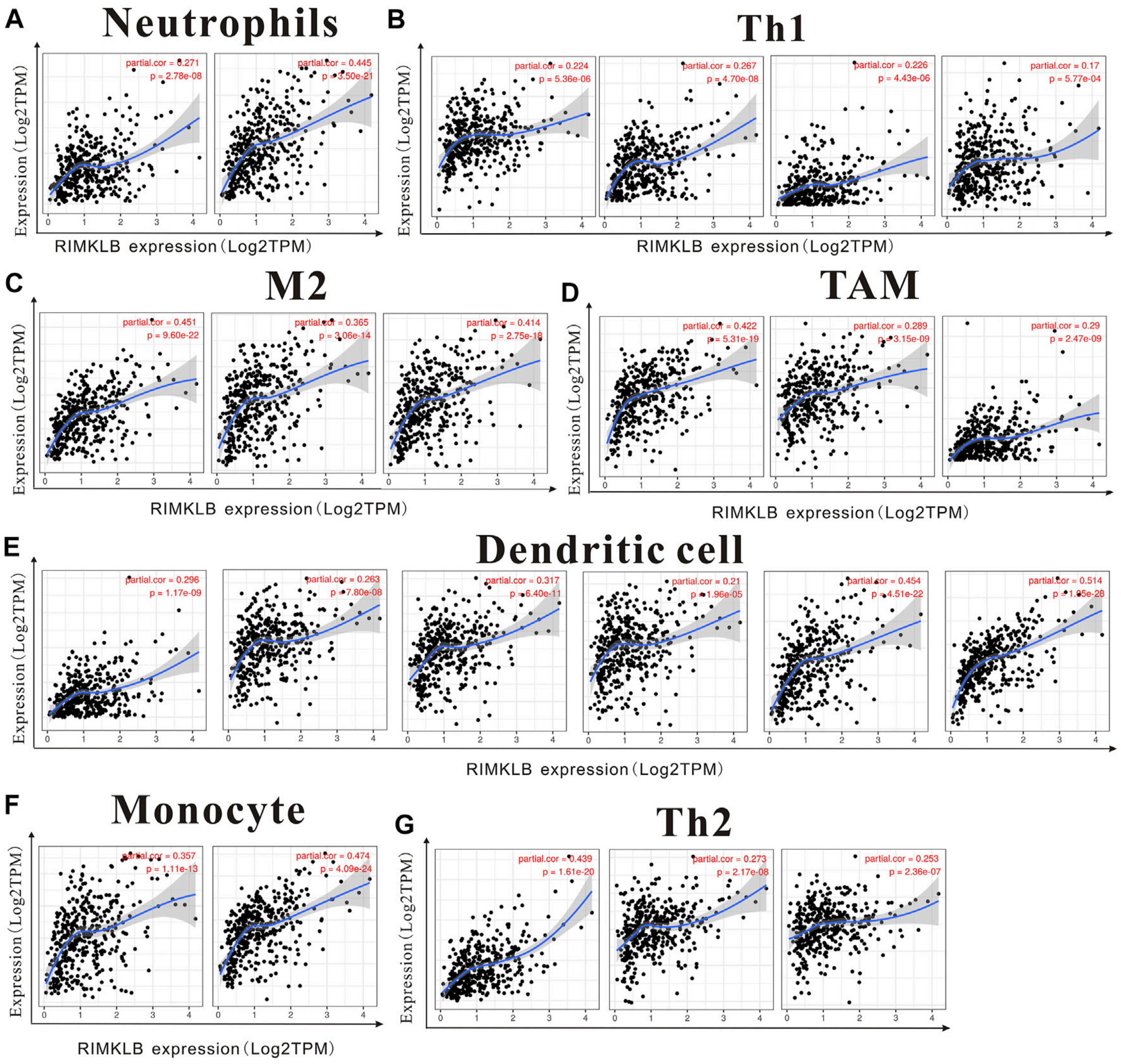


FIGURE 7 | The expression of RIMKLB and the correlation with immune infiltration in COAD (colon adenocarcinoma). Markers include ITGAM, and CCR7 of Neutrophils; TBX21, STAT1, STAT4, and TNF of Th1; CD86 and CSF1R of monocytes; CD163, and VSIG4 of M2 macrophages; CCL2, CD68, and IL10 of TAMs (tumor-associated macrophages); HLA-DPB1, HLA-DRA, HLA-DPA1, CD1C, NRP, and ITGAX of Dendritic cell; GATA3, GATA6, and GATA5A of Th2. **(A–G)** Scatterplots of correlations between RIMKLB expression and gene markers of Neutrophils **(A)**, Th1 **(B)**, and M2 macrophages **(C)**, TAMs **(D)**, Dendritic cell **(E)**, monocytes **(F)**, and Th2 in COAD **(G)**.

Schematic representation for functional relevance of RIMKLB gene in the oncogenesis of colorectal cancer and its candidature as a correlation with immune cells and biological pathways is in **Figure 9**.

As far as we know, there are very few studies on RIMKLB and Immune infiltration at present. This is the first study to find a close correlation between RIMKLB and immune infiltration in CRC. TIMER analysis showed that the mRNA level of RIMKLB was closely related to TIICs, including CD4$^+$ and CD8$^+$ T cells, B cells, TAMs M1 and M2 macrophages, neutrophils, monocytes, natural killer cells, dendritic cells, Th1, Th2, Tfh, and Th17. In addition, the expression of RIMKLB expression was significantly correlated with the infiltration level of immune checkpoint inhibitors (ICIs), and enrichment analysis showed that RIMKLB was positively correlated with immunoinflammatory pathways. This study explored the correlation between RIMKLB and the immune microenvironment, providing new ideas and targets for CRC immunotherapy.

Our study found that there were differences in the expression of RIMKLB between COAD and READ. High expression of RIMKLB in rectal cancer indicated poor OS and DFS, while there was no significant statistical correlation in the case of colon cancer. In COAD and READ, the expression of RIMKLB maintained a high degree of consistency with tumor immune cell infiltration and the expression of immune examination, except for B cell infiltration and PD1 expression. For this reason, we first examined the differences in sample size between COAD and READ groups, and secondly, the possible differences in the pathogenesis of the two cancer types. A retrospective analysis comparing right-sided colon cancer (RCC), left-sided colon cancer (LCC), and colorectal cancer with regards to tumor

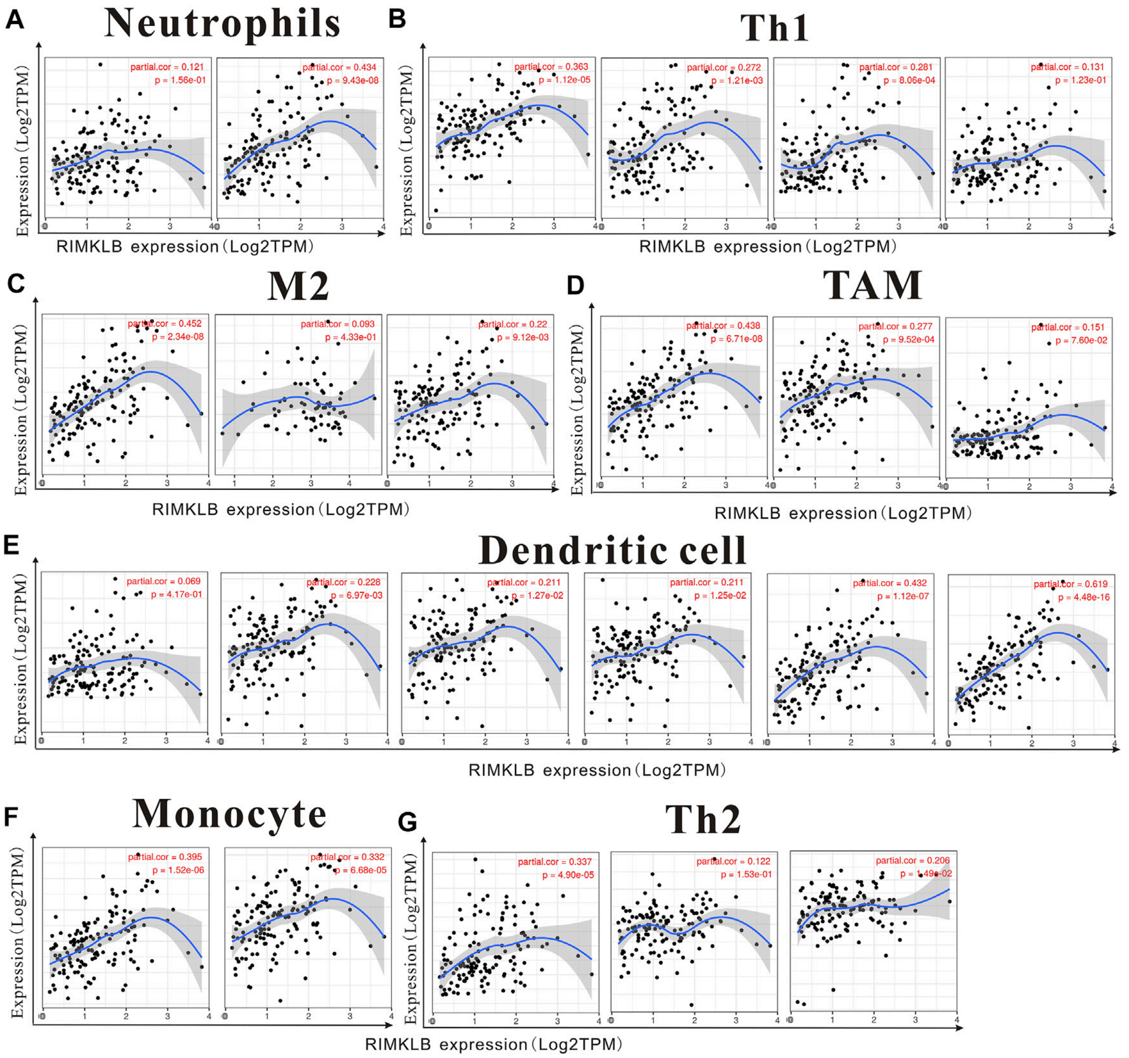


FIGURE 8 | The expression of RIMKLB and the correlation with immune infiltration in READ (recutm adenocarcinoma). Markers include ITGAM, and CCR7 of Neutrophils; TBX21, STAT1, STAT4, and TNF of Th1; CD86 and CSF1R of monocytes; CD163, and VSIG4 of M2 macrophages; CCL2, CD68, and IL10 of TAMs (tumor-associated macrophages); HLA-DPB1, HLA-DRA, HLA-DPA1, CD1C, NRP, and ITGAX of Dendritic cell; GATA3, GATA6, and GATA5A of Th2. **(A–G)** Scatterplots of correlations between RIMKLB expression and gene markers of Neutrophils **(A)**, Th1 **(B)**, and M2 macrophages **(C)**, TAMs **(D)**, Dendritic cell **(E)**, monocytes **(F)**, and Th2 in COAD.

status, differentiation degree, infiltration depth and diameter showed that TNM staging and PFS of RCC was lower than that of the LCC and rectal cancer; hence, survival may be associated with inherent position characteristics (Gao et al., 2017). Studies have found that in addition to anatomical differences, RCC and LCC are also different in embryo origin and metastasis patterns and drug target composition (Tamas et al., 2015). Human colon and rectal cancers were comprehensively analyzed by the Cancer Genome Map Network to identify possible genetic differences between them. Research data show that non-high-mutation tumors correspond to CIN phenotype, while high-mutation tumors correspond to microsatellite instability (MSI) phenotype (Network, 2012). Studies have identified common tumor-initiating events involving APC, KRAS, and TP53 genes in RCC, LCC, and rectal cancer through comparative somatic and proteomic analyses of the three cancer types. However, The sequence of each event in tumor development and selection of downstream somatic changes is different at all three anatomic sites, which may have therapeutic relevance in these highly complex and heterogeneous tumors (Imperial et al., 2018). Therefore, in our study, the slight differences between the two may be closely related to the above reasons, which need further research and verification.

Immunotherapy is a new type of cancer treatment. The strategy is to use the patient's own immune system to fight cancer cells. Tumor immunotherapy overcomes the major problem of specificity in chemotherapy and radiotherapy. Although immunotherapy has dramatically changed the treatment outlook for many advanced cancers, the benefits of CRC to date have been limited to patients with high microsatellite instability (MSI-H) DNA mismatched repair defect (dMMR) tumors, and several randomized controlled trials are under

TABLE 3 | TOP genes positively correlated with RIMKLB in CRC.

Correlated gene	Cytoband	Spearman's correlation	*p*-Value	q-Value
AKT3	1q43-q44	0.706	2.12E-80	4.21E-76
MPDZ	9p23	0.689	4.35E-75	4.32E-71
PKD2	4q22.1	0.681	1.05E-72	6.96E-69
MAP1B	5q13.2	0.679	5.94E-72	2.95E-68
LHFPL6	13q13.3-q14.11	0.674	1.09E-70	4.35E-67
MEIS1	2p14	0.673	1.86E-70	5.52E-67
DNAAF9	20p13	0.673	1.95E-70	5.52E-67
BNC2	9p22.3-p22.2	0.669	2.57E-69	6.39E-66
TNS1	2q35	0.665	4.52E-68	9.97E-65
SLIT2	4p15.31	0.664	5.52E-68	1.10E-64
FBXL7	5p15.1	0.664	6.19E-68	1.12E-64
AMOTL1	11q21	0.663	1.28E-67	2.12E-64
DZIP1	13q32.1	0.662	2.38E-67	3.64E-64
HEG1	3q21.2	0.661	3.57E-67	5.07E-64
ARHGEF25	12q13.3	0.661	5.62E-67	7.44E-64
PTPRM	18p11.23	0.659	1.81E-66	2.25E-63
ZEB1	10p11.22	0.655	1.47E-65	1.72E-62
FILIP1	6q14.1	0.653	6.50E-65	7.17E-62
MCC	5q22.2	0.652	1.24E-64	1.30E-61
JAM3	11q25	0.651	1.69E-64	1.68E-61
TUB	11p15.4	0.648	9.10E-64	8.61E-61
STON1	2p16.3	0.647	1.75E-63	1.58E-60
WHAMMP2	15q13.1	0.646	2.66E-63	2.30E-60
JCAD	10p11.23	0.646	2.85E-63	2.36E-60
SALL2	14q11.2	0.646	3.71E-63	2.95E-60

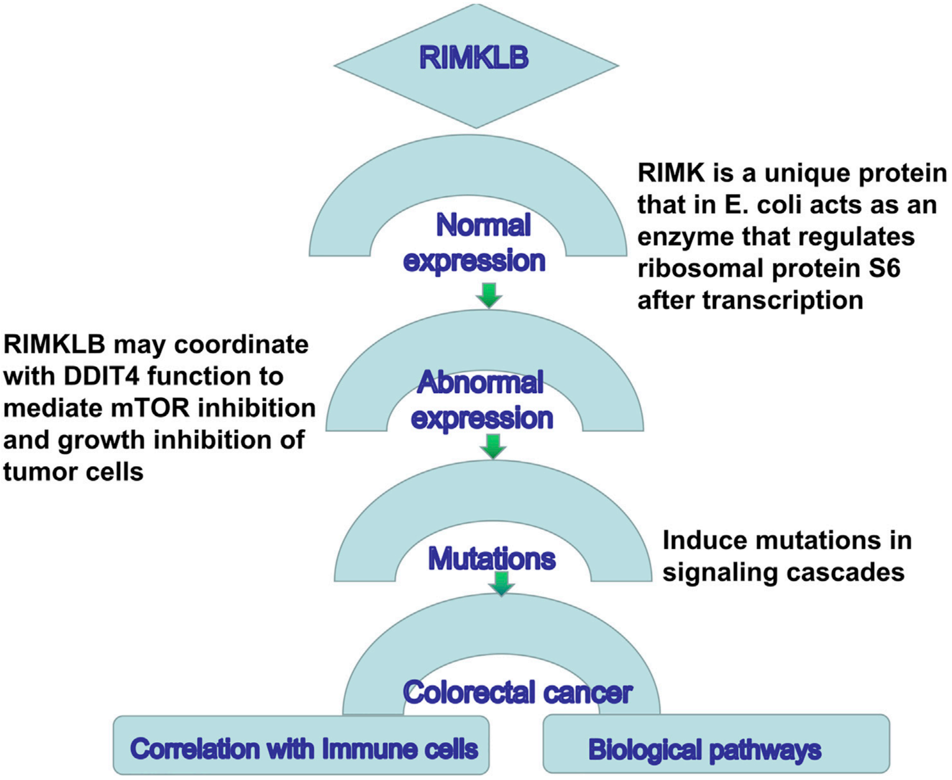

RIMKLB
Normal expression
RIMK is a unique protein that in E. coli acts as an enzyme that regulates ribosomal protein S6 after transcription
RIMKLB may coordinate with DDIT4 function to mediate mTOR inhibition and growth inhibition of tumor cells
Abnormal expression
Mutations
Induce mutations in signaling cascades
Colorectal cancer
Correlation with Immune cells
Biological pathways

FIGURE 9 | Schematic representation for functional relevance of RIMKLB gene in the oncogenesis of colorectal cancer and its candidature as a correlation with immune cells and biological pathways.

way to move immunotherapy to first-line and adjuvant therapy for metastatic cancers (Franke et al., 2019). In recent years, a lot of work has been done to evaluate the prognostic value of various immune cell subsets. In general, cytotoxic T cells, memory T cells, Th1 cells, Tfh cells and B cells are associated with prolonged survival, while increased density of Treg cells, myeloid-derived suppressor cells and neutrophils is associated with poor prognosis (Bruni et al., 2020). Similar results were found in our study. In our study, we found that in colon and rectal cancer, the expression level of RIMKLB was significantly correlated with most immune marker groups of various immune cells and different T cells. Interestingly, we found that neutrophils, Th1, M2 macrophages, TAM, DCs, monocytes, and Th2 were strongly correlated with the expression of RIMKLB expression in the colon and rectum. ICIs is used to target and/or block immune checkpoint protein ligands on the surface of T cells or other immune cell subsets in order to restore immune function. However, the high activation and overexpression of immune checkpoints in cancer lead to the suppression of anti-tumor immune response, which is conducive to the proliferation and diffusion of malignant cells (Pardoll, 2015; Gonzalez et al., 2018). ICIs, specifically PD-1, PDL-1 and CTLA-4 inhibitors, have been approved for the treatment of a variety of solid tumors. Pd-1 and CTLA-4 are both negative costimulatory molecules, and when inhibited, they enhance the activation of T cells and eventually kill tumor cells (Wei et al., 2018). ICIs can be used for tumors with MSI-H and high tumor mutational burden (TMB) in chemo-resistant environments. The most important biomarkers that should be routinely examined in clinical practice include PDL-1, MSI and TMB (Spencer et al., 2016; Mazloom et al., 2020). Our study found that the expression of RIMKLB was significantly correlated with ICIs, specifically with the infiltration levels of PD1, PD-L1, and CTLA4. At the same time, the enrichment analysis of GO pathway suggested that this gene was also involved in immune function. Taken together, these findings suggest that RIMKLB may be closely related to CRC immunotherapy, although further verification is needed.

The role of PI3K-Akt signaling pathway in the occurrence and progression of CRC and its important role in drug resistance have been reported earlier (Narayanankutty, 2019). Studies have confirmed that overexpression of IMPDH2 can promote cell G1/S phase cycle transition by activating the PI3K/AKT/mTOR and PI3K/AKT/FOXO1 pathways, and promote cell invasion, migration and EMT by regulating the PI3K/AKT/mTOR pathway (Narayanankutty, 2019). Han et al. (2020) found that loss of MLH1 reduced CTX sensitivity through HER-2/PI3K/AKT signal transduction and anti-apoptosis and induced activation of HER-2/PI3K/AKT signaling pathway, leading to cetuximab resistance in colon cancer. FAT4 can partially regulate PI3K activity to promote autophagy and inhibit EMT through PI3K/AKT/mTOR and PI3K/AKT/GSK-3β signaling pathways (Wei et al., 2019). Patra et al. (2021) found that COL11A1 plays an important role in regulating cell division, differentiation, proliferation, migration, growth and apoptosis of intestinal and colon cells, and it can disrupt a variety of signaling pathways that affect tumor development, such as RTK-RAS-PI3K, Wnt, TGF-β_2 and TP53 pathways. At present, most studies have confirmed that the PI3K-Akt signaling pathway is regulated by multiple factors and plays a role in the occurrence, development and treatment of tumors. In our study, we found that RIMKLB is enriched in the PI3K-Akt pathway, suggesting that this molecule plays a role in CRC progression or treatment, but the specific mechanism needs further experimental verification.

Our research has its limitations. First of all, our study lacks cytological and animal experiments, and the specific mechanism is not clear. Further molecular cytological studies are needed in the future. Second, our retrospective study and small sample size failed to obtain immunotherapy data for these patients; Finally, there is a lack of data on molecular indicators (such as MSI, TP53, and TMB, etc) associated with colorectal cancer prognosis and immunotherapy, so further improvement is needed.

CONCLUSION

Our study revealed the relationship between RIMKLB and the prognosis of CRC for the first time, and also found that this molecule was closely related to the invasion of CRC immune cells and ICIs. Thus, our study provides an important basis for the immunotherapy of CRC, the mechanism of immune resistance, and the identification of new immune-related therapeutic targets.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

KW, BZ and KC conceived and designed the study. YC, SD, LY, JG, FM, YX, LQ, ZJ, CZ, WC, XN, HL, ZS, FS, KT, and JW, collected and analysed data. YC, SD and LY wrote the paper. KW, BZ and KC reviewed and edited the manuscript. All authors read and approved the manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (No. 82170678) and Hubei Province Key Research and Development Program of China (Science and Technology Innovation Special Project No. 2021BAA044), and Wuhan Strong Magnetic Field Interdisciplinary Fund (No. WHMF202113), and free Innovation Pre-Research Fund and Platform Scientific Research Fund in 2020 (F016.02004.20001.023).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.818994/full#supplementary-material

REFERENCES

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., et al. (2013). Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the Immune Landscape in Human Cancer. *Immunity* 39, 782–795. doi:10.1016/j.immuni.2013.10.003

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer J. Clinicians* 68, 394–424. doi:10.3322/caac.21492

Bruni, D., Angell, H. K., and Galon, J. (2020). The Immune Contexture and Immunoscore in Cancer Prognosis and Therapeutic Efficacy. *Nat. Rev. Cancer* 20, 662–680. doi:10.1038/s41568-020-0285-7

Chandrashekar, D. S., Bashel, B., Balasubramanya, S. A. H., Creighton, C. J., Ponce-Rodriguez, I., Chakravarthi, B. V. S. K., et al. (2017). UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia* 19, 649–658. doi:10.1016/j.neo.2017.05.002

Collard, F., Stroobant, V., Lamosa, P., Kapanda, C. N., Lambert, D. M., Muccioli, G. G., et al. (2010). Molecular Identification of N-Acetylaspartylglutamate Synthase and β-Citrylglutamate Synthase. *J. Biol. Chem.* 285, 29826–29833. doi:10.1074/jbc.m110.152629

Dekker, E., Tanis, P. J., Vleugels, J. L. A., Kasi, P. M., and Wallace, M. B. (2019). Colorectal Cancer. *The Lancet* 394, 1467–1480. doi:10.1016/s0140-6736(19)32319-0

Duan, S., Huang, W., Liu, X., Liu, X., Chen, N., Xu, Q., et al. (2018). IMPDH2 Promotes Colorectal Cancer Progression through Activation of the PI3K/AKT/mTOR and PI3K/AKT/FOXO1 Signaling Pathways. *J. Exp. Clin. Cancer Res.* 37, 304. doi:10.1186/s13046-018-0980-3

Franke, A. J., Skelton, W. P., Starr, J. S., Parekh, H., Lee, J. J., Overman, M. J., et al. (2019). Immunotherapy for Colorectal Cancer: A Review of Current and Novel Therapeutic Approaches. *J. Natl. Cancer Inst.* 111, 1131–1141. doi:10.1093/jnci/djz093

Galon, J., Fridman, W.-H., and Pagès, F. (2007). The Adaptive Immunologic Microenvironment in Colorectal Cancer: A Novel Perspective: Figure 1. *Cancer Res.* 67, 1883–1886. doi:10.1158/0008-5472.can-06-4806

Ganesh, K., Stadler, Z. K., Cercek, A., Mendelsohn, R. B., Shia, J., Segal, N. H., et al. (2019). Immunotherapy in Colorectal Cancer: Rationale, Challenges and Potential. *Nat. Rev. Gastroenterol. Hepatol.* 16, 361–375. doi:10.1038/s41575-019-0126-x

Gao, X. H., Yu, G. Y., Gong, H. F., Liu, L. J., Xu, Y., Hao, L. Q., et al. (2017). Differences of Protein Expression Profiles, KRAS and BRAF Mutation, and Prognosis in Right-Sided colon, Left-Sided colon and Rectal Cancer. *Sci. Rep.* 7, 7882. doi:10.1038/s41598-017-08413-z

Gonzalez, H., Hagerling, C., and Werb, Z. (2018). Roles of the Immune System in Cancer: from Tumor Initiation to Metastatic Progression. *Genes Dev.* 32, 1267–1284. doi:10.1101/gad.314617.118

Grenga, L., Little, R. H., and Malone, J. G. (2017). Quick Change: post-transcriptional Regulation in Pseudomonas. *Fems Microbiol. Lett.* 364, fnx125. doi:10.1093/femsle/fnx125

Gulubova, M. V., Ananiev, J. R., Vlaykova, T. I., Yovchev, Y., Tsoneva, V., and Manolova, I. M. (2012). Role of Dendritic Cells in Progression and Clinical Outcome of colon Cancer. *Int. J. Colorectal Dis.* 27, 159–169. doi:10.1007/s00384-011-1334-1

Han, Y., Peng, Y., Fu, Y., Cai, C., Guo, C., Liu, S., et al. (2020). MLH1 Deficiency Induces Cetuximab Resistance in Colon Cancer *via* Her-2/PI3K/AKT Signaling. *Adv. Sci.* 7, 2000112. doi:10.1002/advs.202000112

Huang, C., Zhao, J., and Zhu, Z. (2021). Prognostic Nomogram of Prognosis-Related Genes and Clinicopathological Characteristics to Predict the 5-Year Survival Rate of Colon Cancer Patients. *Front. Surg.* 8, 681721. doi:10.3389/fsurg.2021.681721

Imperial, R., Ahmed, Z., Toor, O. M., Erdoğan, C., Khaliq, A., Case, P., et al. (2018). Comparative Proteogenomic Analysis of Right-Sided colon Cancer, Left-Sided colon Cancer and Rectal Cancer Reveals Distinct Mutational Profiles. *Mol. Cancer* 17, 177. doi:10.1186/s12943-018-0923-9

Japanese Gastric Cancer Association (2017). Japanese Gastric Cancer Treatment Guidelines 2014 (Ver. 4). *Gastric Cancer* 20, 1–19. doi:10.1007/s10120-016-0622-4

Kino, K., Arai, T., and Arimura, Y. (2011). Poly-α-Glutamic Acid Synthesis Using a Novel Catalytic Activity of RimK from Escherichia coli K-12. *Appl. Environ. Microbiol.* 77, 2019–2025. doi:10.1128/aem.02043-10

Li, B., Severson, E., Pignon, J.-C., Zhao, H., Li, T., Novak, J., et al. (2016). Comprehensive Analyses of Tumor Immunity: Implications for Cancer Immunotherapy. *Genome Biol.* 17, 174. doi:10.1186/s13059-016-1028-7

Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. S., et al. (2017). TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* 77, e108–e110. doi:10.1158/0008-5472.can-17-0307

Maekura, K., Tsukamoto, S., Hamada-Kanazawa, M., and Takano, M. (2021). Rimklb Mutation Causes Male Infertility in Mice. *Sci. Rep.* 11, 4604. doi:10.1038/s41598-021-84105-z

Mauri, G., Sartore-Bianchi, A., Russo, A. G., Marsoni, S., Bardelli, A., and Siena, S. (2019). Early-onset Colorectal Cancer in Young Individuals. *Mol. Oncol.* 13, 109–131. doi:10.1002/1878-0261.12417

Mazloom, A., Ghalehsari, N., Gazivoda, V., Nimkar, N., Paul, S., Gregos, P., et al. (2020). Role of Immune Checkpoint Inhibitors in Gastrointestinal Malignancies. *J. Clin. Med.* 9. doi:10.3390/jcm9082533

Mizuno, H., Kitada, K., Nakai, K., and Sarai, A. (2009). PrognoScan: a New Database for Meta-Analysis of the Prognostic Value of Genes. *BMC Med. Genomics* 2, 18. doi:10.1186/1755-8794-2-18

Narayanankutty, A. (2019). PI3K/Akt/mTOR Pathway as a Therapeutic Target for Colorectal Cancer: A Review of Preclinical and Clinical Evidence. *Cdt* 20, 1217–1226. doi:10.2174/1389450120666190618123846

Network, C. G. A. (2012). Comprehensive Molecular Characterization of Human colon and Rectal Cancer. *Nature* 487, 330–337. doi:10.1038/nature11252

Ohtani, H. (2007). Focus on TILs: Prognostic Significance of Tumor Infiltrating Lymphocytes in Human Colorectal Cancer. *Cancer Immun.* 7, 4.

Pardoll, D. (2015). Cancer and the Immune System: Basic Concepts and Targets for Intervention. *Semin. Oncol.* 42, 523–538. doi:10.1053/j.seminoncol.2015.05.003

Patra, R., Das, N. C., and Mukherjee, S. (2021). Exploring the Differential Expression and Prognostic Significance of the COL11A1 Gene in Human Colorectal Carcinoma: An Integrated Bioinformatics Approach. *Front. Genet.* 12, 608313. doi:10.3389/fgene.2021.608313

Pletnev, P. I., Nesterchuk, M. V., Rubtsova, M. P., Serebryakova, M. V., Dmitrieva, K., Osterman, I. A., et al. (2019). Oligoglutamylation of E. coli Ribosomal Protein S6 Is under Growth Phase Control. *Biochimie* 167, 61–67. doi:10.1016/j.biochi.2019.09.008

Rahma, O. E., and Hodi, F. S. (2019). The Intersection between Tumor Angiogenesis and Immune Suppression. *Clin. Cancer Res.* 25, 5449–5457. doi:10.1158/1078-0432.ccr-18-1543

Samstein, R. M., Lee, C.-H., Shoushtari, A. N., Hellmann, M. D., Shen, R., Janjigian, Y. Y., et al. (2019). Tumor Mutational Load Predicts Survival after Immunotherapy across Multiple Cancer Types. *Nat. Genet.* 51, 202–206. doi:10.1038/s41588-018-0312-8

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer Statistics, 2021. *CA A. Cancer J. Clin.* 71, 7–33. doi:10.3322/caac.21654

Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer Statistics, 2019. *CA A. Cancer J. Clin.* 69, 7–34. doi:10.3322/caac.21551

Spencer, K. R., Wang, J., Silk, A. W., Ganesan, S., Kaufman, H. L., and Mehnert, J. M. (2016). Biomarkers for Immunotherapy: Current Developments and Challenges. *Am. Soc. Clin. Oncol. Educ. Book* 35, e493–e503. doi:10.1200/edbk_160766

Tada, K., Kitano, S., Shoji, H., Nishimura, T., Shimada, Y., Nagashima, K., et al. (2016). Pretreatment Immune Status Correlates with Progression-free Survival in Chemotherapy-Treated Metastatic Colorectal Cancer Patients. *Cancer Immunol. Res.* 4, 592–599. doi:10.1158/2326-6066.cir-15-0298

Tamas, K., Walenkamp, A. M. E., de Vries, E. G. E., van Vugt, M. A. T. M., Beets-Tan, R. G., van Etten, B., et al. (2015). Rectal and colon Cancer: Not Just a Different Anatomic Site. *Cancer Treat. Rev.* 41, 671–679. doi:10.1016/j.ctrv.2015.06.007

Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). GEPIA: a Web Server for Cancer and normal Gene Expression Profiling and Interactive Analyses. *Nucleic Acids Res.* 45, W98–W102. doi:10.1093/nar/gkx247

Wang, Y., Han, E., Xing, Q., Yan, J., Arrington, A., Wang, C., et al. (2015). Baicalein Upregulates DDIT4 Expression Which Mediates mTOR Inhibition and Growth Inhibition in Cancer Cells. *Cancer Lett.* 358, 170–179. doi:10.1016/j.canlet.2014.12.033

Wei, R., Xiao, Y., Song, Y., Yuan, H., Luo, J., and Xu, W. (2019). FAT4 Regulates the EMT and Autophagy in Colorectal Cancer Cells in Part *via* the PI3K-AKT Signaling axis. *J. Exp. Clin. Cancer Res.* 38, 112. doi:10.1186/s13046-019-1043-0

Wei, S. C., Duffy, C. R., and Allison, J. P. (2018). Fundamental Mechanisms of Immune Checkpoint Blockade Therapy. *Cancer Discov.* 8, 1069–1086. doi:10.1158/2159-8290.cd-18-0367

Wu, X., Qu, D., Weygant, N., Peng, J., and Houchen, C. W. (2020). Cancer Stem Cell Marker DCLK1 Correlates with Tumorigenic Immune Infiltrates in the Colon and Gastric Adenocarcinoma Microenvironments. *Cancers (Basel)* 12, 12. doi:10.3390/cancers12020274

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

ORIGINAL RESEARCH
published: 20 April 2022
doi: 10.3389/fimmu.2022.870531

Accurate Machine Learning Model to Diagnose Chronic Autoimmune Diseases Utilizing Information From B Cells and Monocytes

Yuanchen Ma[1†], *Jieying Chen*[1†], *Tao Wang*[1], *Liting Zhang*[1], *Xinhao Xu*[2], *Yuxuan Qiu*[2], *Andy Peng Xiang*[1] *and Weijun Huang*[1*]

[1] Center for Stem Cell Biology and Tissue Engineering, Key Laboratory for Stem Cells and Tissue Engineering, Ministry of Education, Sun Yat-sen University, Guangzhou, China, [2] Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China

OPEN ACCESS

Edited by:
Xu-jie Zhou,
Peking University First Hospital, China

Reviewed by:
Altuna Akalin,
Helmholtz Association of German Research Centers (HZ), Germany
Shibiao Wan,
St. Jude Children's Research Hospital, United States

***Correspondence:**
Weijun Huang
hweijun@mail.sysu.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Autoimmune and Autoinflammatory Disorders, a section of the journal Frontiers in Immunology

Received: *07 February 2022*
Accepted: *22 March 2022*
Published: *20 April 2022*

Citation:
Ma Y, Chen J, Wang T, Zhang L, Xu X, Qiu Y, Xiang AP and Huang W (2022) Accurate Machine Learning Model to Diagnose Chronic Autoimmune Diseases Utilizing Information From B Cells and Monocytes. Front. Immunol. 13:870531. doi: 10.3389/fimmu.2022.870531

Heterogeneity and limited comprehension of chronic autoimmune disease pathophysiology cause accurate diagnosis a challenging process. With the increasing resources of single-cell sequencing data, a reasonable way could be found to address this issue. In our study, with the use of large-scale public single-cell RNA sequencing (scRNA-seq) data, analysis of dataset integration (3.1×10^5 PBMCs from fifteen SLE patients and eight healthy donors) and cellular cross talking (3.8×10^5 PBMCs from twenty-eight SLE patients and eight healthy donors) were performed to identify the most crucial information characterizing SLE. Our findings revealed that the interactions among the PBMC subpopulations of SLE patients may be weakened under the inflammatory microenvironment, which could result in abnormal emergences or variations in signaling patterns within PBMCs. In particular, the alterations of B cells and monocytes may be the most significant findings. Utilizing this powerful information, an efficient mathematical model of unbiased random forest machine learning was established to distinguish SLE patients from healthy donors *via* not only scRNA-seq data but also bulk RNA-seq data. Surprisingly, our mathematical model could also accurately identify patients with rheumatoid arthritis and multiple sclerosis, not just SLE, *via* bulk RNA-seq data (derived from 688 samples). Since the variations in PBMCs should predate the clinical manifestations of these diseases, our machine learning model may be feasible to develop into an efficient tool for accurate diagnosis of chronic autoimmune diseases.

Keywords: chronic autoimmune disease, accurate diagnosis, machine learning (ML), scRNA-seq, cellular cross talking

INTRODUCTION

Systemic lupus erythematosus (SLE), multiple sclerosis (MS), and rheumatoid arthritis (RA) are all chronic autoimmune diseases associated with progressive widespread organ damage (1–3). The course of these three diseases is typically progressive with intermittent remission (4, 5). It is generally accepted that early treatment could increase the remission probability of these diseases and improve their prognosis (6, 7). If appropriate treatment is not given in a timely manner, these diseases may progress, causing work disability and life quality reduction for patients. Furthermore, such progression would lead to enormous financial burdens to the patients, their families, and society (8–10). Hence, it is crucial to develop an efficient method of accurate diagnosis to enable early intervention for these diseases.

Unfortunately, it seems that diagnosing SLE, MS, and RA may still be a challenging process that relies on a set of criteria (11–13), including clinical manifestations, functional outcomes, and serological and radiological evidence, that have to be met to make an accurate diagnosis (14, 15). Under non-specific and insensitive criteria, the misdiagnosis and underdiagnosis of these diseases are relatively common (16). The average time from symptom onset to diagnosis confirmation was approximately two years (17). This may cause patients to miss the optimal time for treatment. To break the bottleneck of early diagnosis, many studies have focused on biomarker detection to develop an accurate diagnostic criterion (18–21). However, the results were unsatisfying, owing to the tremendous heterogeneity of these diseases and limited comprehension of the disease pathophysiology (22).

In detail, although it is well known that the loss of immune tolerance and persistent release of autoantibodies are the two important bases for the pathophysiology of chronic autoimmune disease (23, 24), most studies have focused on investigating the contribution of certain cellular or molecular mechanisms rather than comprehensively and systematically illustrating the pathogenesis. This might be due to the limitation of methods or means. With the development of single-cell sequencing technology, the increased resources of data, and the improvement of bioinformatic tools (e.g., Seurat, SHARP, CellChat, etc.) (25–27), these would together help us to comprehend the pathophysiology of these diseases, thus their crucial features would be efficient for being mined. For example, Nehar-Belaid et al. thoroughly analyzed the major cell types among peripheral blood mononuclear cells and revealed an expanded subpopulation that has a specific interferon-stimulated gene (ISG) expression pattern in SLE patients (28). Meena Subramaniam et al. also found that monocytes from SLE patients highly expressed ISGs (29). Both of these studies comprehensively illuminated the cytological changes of SLEs.

According to these public single-cell RNA sequencing (scRNA-seq) data of SLE, we seek for a feasible way for SLE accurate diagnosis. Firstly, integration and cellular cross-talking analysis were performed to obtain the powerful information labeling the disease. This information was then combined with an unbiased random forestry machine learning algorithm which rendered an efficient mathematical model for SLE diagnosis. The accuracy of the mathematical model to identify patients with RA and MS was also validated. Furthermore, the diagnostic precision of our model was evaluated using an independent SLE cohort (**Figure 1**).

MATERIAL AND METHODS

Data Availability

The single-cell RNA sequencing data were deposited in the Gene Expression Omnibus (GEO), and the accession numbers were GSE137029 and GSE135779 for SLE patients and GSE164378 for healthy donors. Bulk RNA-sequencing data were deposited to GSE72509 and GSE164457 for peripheral blood mononuclear cells (PBMCs) of SLE patients, GSE90081 for PBMCs of RA patients, GSE89408 for synovial tissues of RA patients, GSE159225 for PBMCs of MS patients, and GSE89408 for CD14-positive cells of MS patients, and GSE183204 and GSE169687 for PBMCs of healthy donors.

Integration of Single-Cell RNA Sequencing Data

Reciprocal principal component analysis (RPCA)-based integration could effectively detect a state-specific cell cluster and run significantly faster on large datasets. Compared with other integration tools (e.g., BBKNN and LIGER), RPCA could conserve more distinct cell identities when removing batch effect, particularly for the data of immune cells (30). Considering its balancing capability on batch effect removal and biological variance preserving, RPCA would be used for our dataset integration. Before the integration, two lists were created: one containing merged SLE data and the other containing merged healthy data. These two lists were then combined and integrated through Seurat (version 4.0.5) following the guidelines at https://satijalab.org/seurat/articles/integration_rpca.html.

PBMCs and Their Subpopulation Clustering

To discover SLE-dominant cell clusters, PBMCs and their subpopulations were clustered through Seurat (version 4.0.5), respectively. Cell proportions of each cluster were calculated subsequently. For PBMC cell clustering, each cell subcluster was annotated based on a canonical marker. Any cluster that has SLE cells containing more than 75% would be considered as SLE dominant.

Differential Expression Gene Analysis on SLE-Dominant Cell Clusters

Within those PBMC subpopulations (e.g., B cells and monocytes) which contain the SLE-dominated cluster, differential expression gene (DEG) analysis would be applied on all of their cell clusters with Function *FindAllMarkers* embedded in Seurat (version 4.0.5) to find out useful information that mark the SLE state. Top five genes based on their log2 fold change value were selected as the first part of feature input for machine learning. Meanwhile, these DEG functions were annotated through literature search.

Cellular Cross-Talking Analysis

The machine learning model can be optimized with powerful sources of information. Thus, CellChat (version 1.1.3) analysis was performed following the guidelines at https://github.com/sqjin/CellChat. In details, overall interaction, overall signaling pattern, outgoing/incoming signaling pattern, and ligand–receptor pair were checked step by step. Samples were analyzed independently. Datasets of patients and health donors were analyzed separately and merged to make a comparison analysis. Ligand–receptor pairs which disappeared at SLE were selected as a second part of feature input for machine learning.

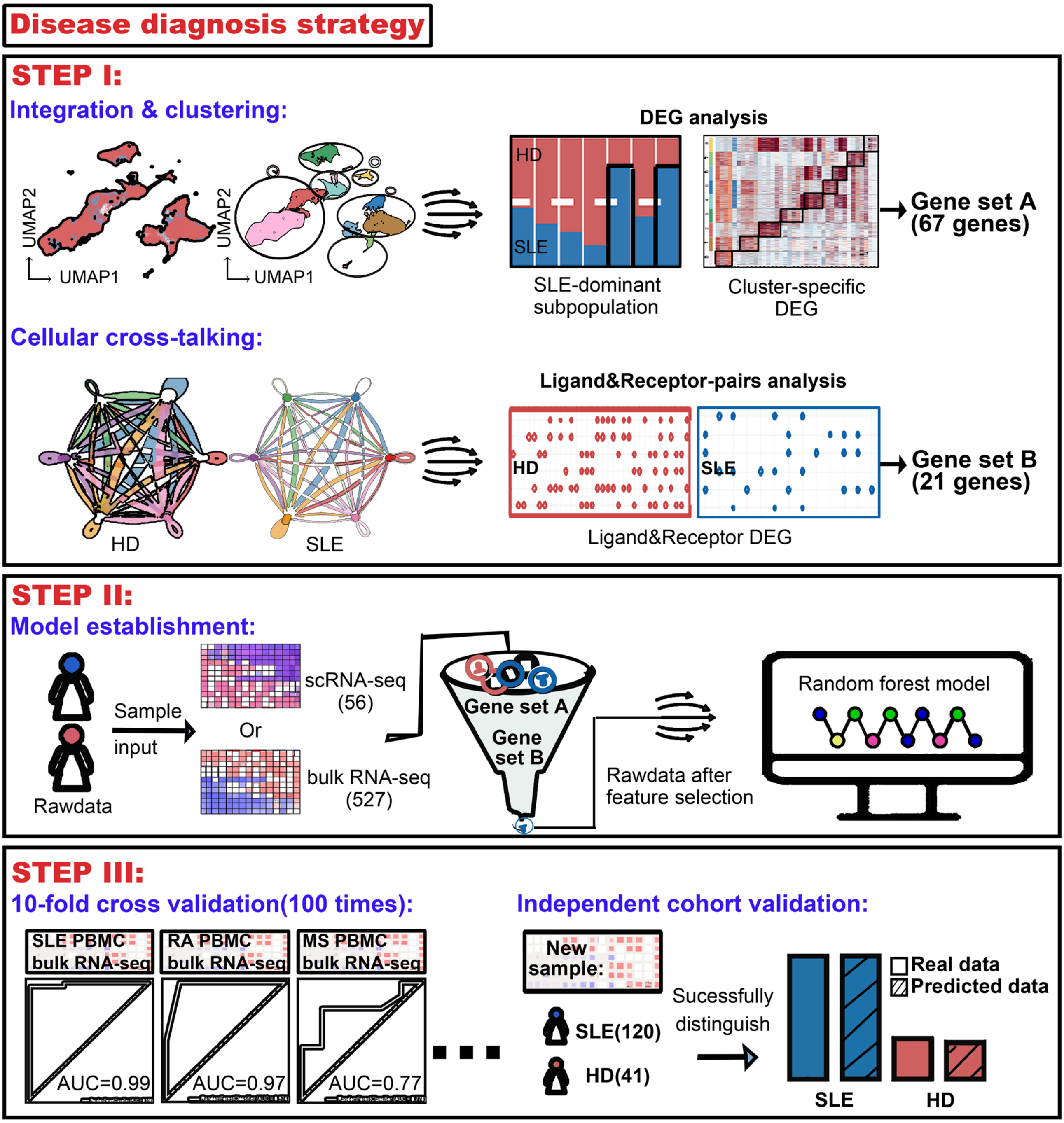


FIGURE 1 | Workflow for establishment of an accurate machine learning model to diagnose chronic autoimmune diseases. STEP I, to figure out the most crucial information that characterizes diseases using public scRNA-seq datasets. From analysis of integration and clustering, 67 top five cluster-specific genes basing on the differential expression gene identification within SLE dominant PBMC subpopulations were derived. From cellular cross-talking analysis, 21 genes constituting ligand–receptor pairs disappeared in SLE patients and showed that more than two kinds of PBMC subpopulation were derived. A union of these two gene sets would be used in the next step. STEP II, to establish the machine learning model diagnosing diseases. A random forest machine learning model was implemented, and genes derived from step I were combined as feature input. 56 and 527 samples were used as sample input for scRNA-seq and bulk RNA-seq data, respectively. STEP III, to validate the accuracy of our machine learning model. Receiver operating characteristic (ROC) analysis was used to test the accuracy, and multiple times of ten-fold cross-validation tests were adopted to avoid bias. The diagnostic accuracy of our model was also validated using an independent bulk RNA-seq cohort containing 120 SLE patients and 41 health donors.

Machine Learning With the Random Forest Model

The random forest machine learning model was implemented with sklearn (version 0.23.2). The gene set which derived from integration and CellChat analysis were combined as feature input, aiming at selecting information within the sequencing datasets, thus improving the performance of the machine learning model. 56 and 527 samples were used as sample input for scRNA-seq and bulk RNA-seq data, respectively. Samples from patients and healthy donors were labeled with 1 and 0, respectively. With the function *train_test_split* within *sklearn.model_selection*, the data were split into two parts, 70% for training and 30% for testing, according to previous study (31). Data balancing was performed when the cell/sample ratio between patients and healthy donors was above 1:2, at random forest model initialization. Receiver operating characteristic (ROC) analysis was used to test models' accuracy. The models for each disease were independent.

To avoid bias of data composition, the sklearn module *StratifiedKFold* was used to split data into ten parts preserving the ratio of samples and perform a ten-fold cross-validation with a loop of one hundred. The average and standard deviation of area under curve (AUC) were documented.

Diagnostic Accuracy Validation of the Machine Learning Model

An independent bulk RNA-seq cohort containing 120 SLE patients and 41 health donors was enrolled into the diagnostic accuracy validation of our machine learning model. Basic information of this cohort including SLE severity, age, and gender was documented. Genes which were used as feature input for the machine learning model were confirmed to be expressed in each sample. The diagnostic accuracy of our machine learning model for SLE and healthy donors was tested separately.

Statistical Analysis

The statistical significance of differential gene expression was analyzed with the Wilcoxon test, a default parameter in function *FindAllMarkers* of Seurat packages.

Software Version

All the software mentioned above were based on R (version 4.1.1) and Python (3.7). Integration analysis and cell clustering were based on Seurat (version 4.0.5), and cellular cross-talking analysis was based on CellChat (1.1.3). Machine learning was based on sklearn (version 0.23.2).

RESULTS

The Limited Alterations of Cell Composition in SLE Patients From the Overall PBMC Perspective

To discover the SLE-dominated alterations of PBMC composition in SLE patients, two single-cell transcriptomic datasets with more than 3.15 × 10^5 cells from 15 SLE patient (GSE137029) and 8 healthy donor (GSE164378) samples were enrolled in our study. The uniform manifold approximation and projection (UMAP) and Louvain algorithm were applied for unsupervised dimension reduction and clustering, respectively (32, 33). As shown in **Figures 2A, B**, the PBMCs of these two datasets could be grouped into sixteen molecularly distinct clusters. The clusters were annotated based on the gene expression values compared to all other cells. The results illustrated two clusters of T cells, B cells, natural killer cells, and erythroid cells, three clusters of monocytes and dendritic cells, and one platelet cluster (**Figures 2A, D**). Unfortunately, SLE-dominated (clusters 13 and 15) clusters were tiny and might come from erythrocytes (HBB specifically expressed). The rest of the cell cluster proportions of SLE patients and healthy donors were evenly balanced or healthy donor dominant (**Figure 2C**). This is partly because the difference between SLE patients and healthy donors might be attenuated under the overall PBMC perspective. Hence, to strengthen the power of detecting SLE-dominant information, further analyses were performed in the subpopulations of PBMCs according to the cluster annotation above.

Identification of SLE-Dominated Clusters in B Cells and Monocytes

Increasing evidence indicates that specialized immune cell subsets are involved in the pathophysiological process of autoimmune diseases through multiplex pathways and signals (34–36). Thus, we re-clustered the subpopulations of PBMCs to identify the SLE-dominated clusters in which the cell proportion of SLE exceeds 75%. Interestingly, the SLE-dominated clusters were identified only in B cells (clusters 2, 6, and 7, **Figures 3A, B**) and monocytes (clusters 1 and 7, **Figures 3E, F**); the rest of the PBMC subpopulation is shown in **Figure S2**. With differential expression gene (DEG) analysis on B cells and monocytes, the top five cluster-specific genes based on their log2 fold change values are shown in **Figures 3C, G**, respectively. All DEG analysis results are shown in **Table S1**. Interferon inflammatory signatures are closely related to the SLE (37). Consistently, we found that cluster 7 of B cells has interferon-stimulated gene (ISG) expression patterns (IFI27, MX1, ISG15, and IFI44L). Moreover, we identified that this cluster simultaneously possess the typical expression patterns of naïve and autoactive B lymphocytes (naïve: IgD+, CD27-, CD38 low, CD24 low; autoactive: TBX21, ITGAX, CXCR5, TRAF5, CR2, **Figure 3D**) (38, 39). In addition, we also found that cluster 1 of monocyte highly expressed ISGs (IFI27, MX1, ISG15, IFI44L), and cluster 7 of monocyte had a proinflammatory character (FKBP5, **Figure 3H**) (40).

Taken together, these findings revealed that there were enhanced signals of an autoreactive/inflammatory state in B cells and monocytes of SLE patients, which suggested the essential roles in the pathophysiological process of SLE.

Weakened Interactions Among the PBMC Subpopulations of SLE Patients

To systematically explore the alterations of PBMCs in SLE patients and obtain a powerful source of information for the training of the machine learning model, we employed CellChat

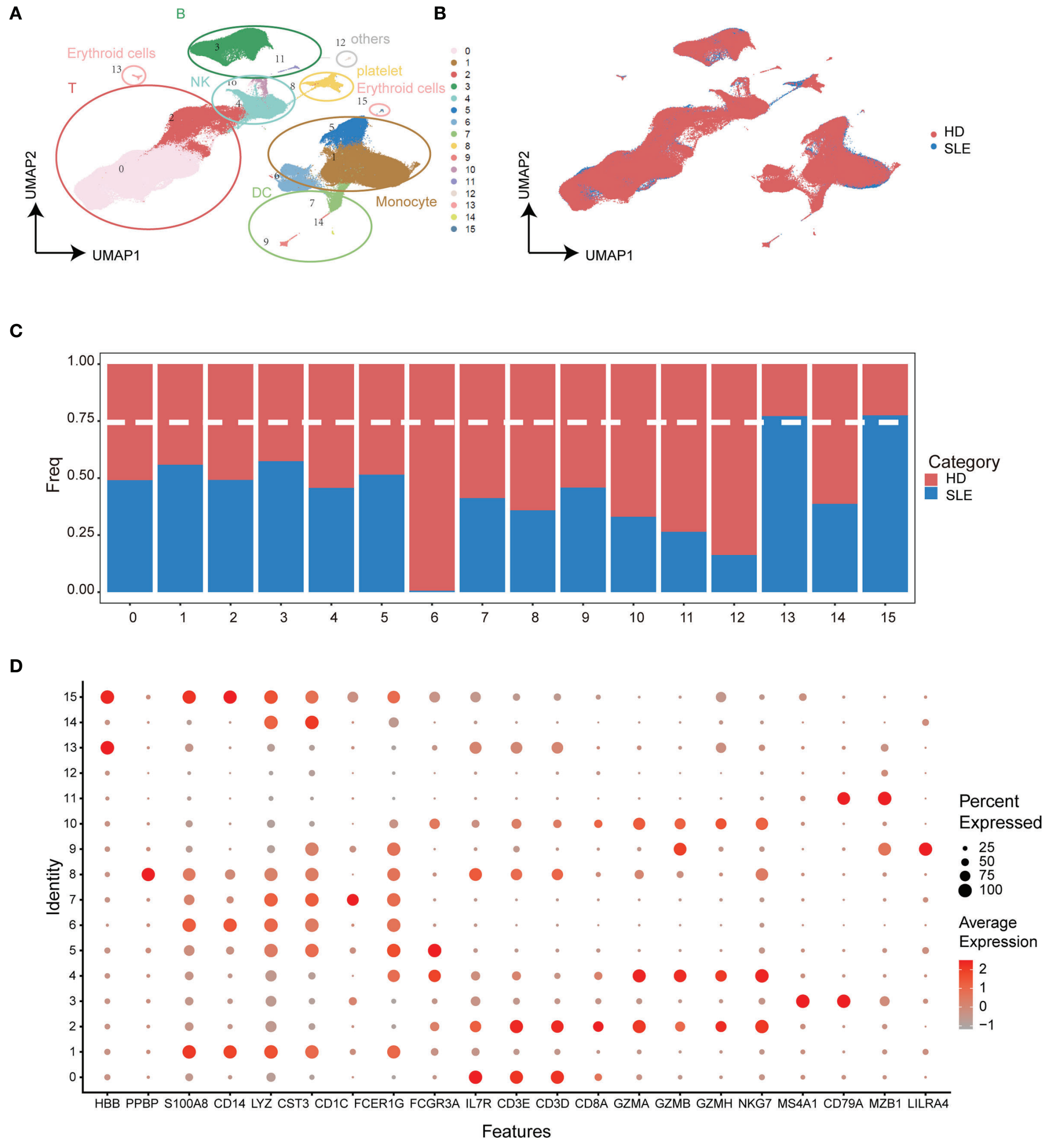


FIGURE 2 | Integration analysis of single-cell RNA sequencing datasets from SLE patients and healthy donors. **(A)** UMAP plot of categorized cell clusters. **(B)** UMAP plot of single-cell PBMCs from fifteen SLE flare patients and eight healthy donors. **(C)** Bar plot of cell proportion in each cell cluster. The dashed line represents the 75% threshold. **(D)** Dot plot of canonical markers for B cells, monocytes, T cells, natural killer cells, dendritic cells, and platelets. The dot size represents the gene (x-axis) percent expression on its corresponding cluster (y-axis). The color represents the average expression of the genes (gray: low, red: high).

to analyze cellular cross talking from scRNA-seq data. Three scRNA-seq datasets (GSE137029, 15 adult patients with SLE; GSE135779, 13 child patients with SLE; GSE164378, 8 healthy donors) with more than 3.80 × 10^5 cells were included in this analysis.

The total number and strength of ligand–receptor pairs were significantly reduced in both adult and child SLE patients compared with healthy donors (**Figure 4A**). Remarkably, the interactions of PBMC subpopulations in SLE patients were weakened (**Figure 4B**). Comparing overall and detailed outgoing/incoming signaling pattern variations among SLE and healthy donors, we identified that abundant signal patterns could be observed for the healthy donors, but in the SLE groups, the number of involved pathways was reduced (**Figures 4C, D**). In detail, there were several signal patterns that specifically disappeared under the disease state. Among them, FLT3, CD48, and TGF-beta signal patterns have been reported to have a negative correlation with SLE development (41–44). Taken together, the disappearance of multiple signal patterns might be a potential feature during SLE development.

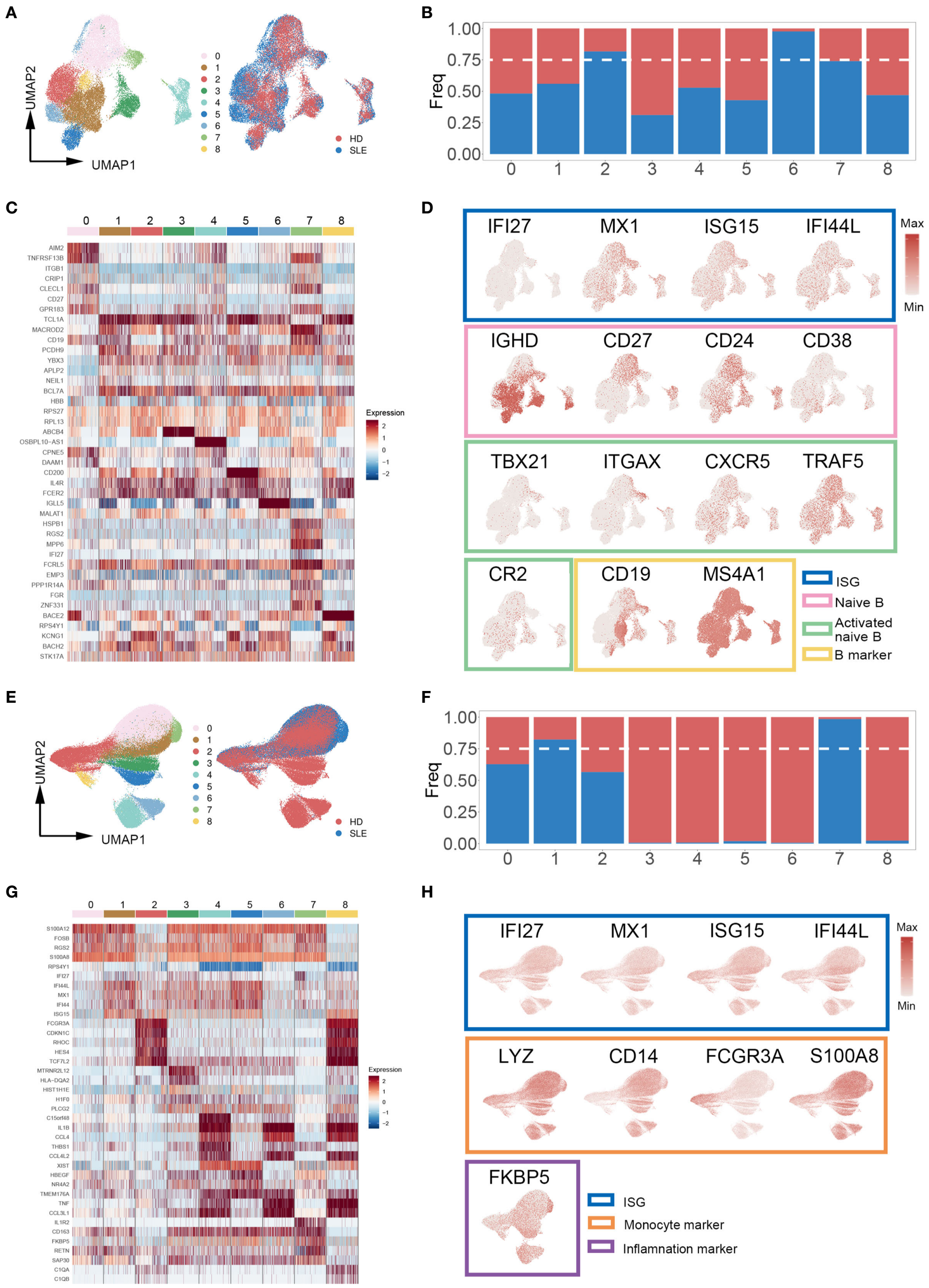


FIGURE 3 | Cell proportion analysis of re-clustered B cells and monocytes. **(A, E)** UMAP plot of re-clustered B cells and monocytes from SLE patients and healthy donors, respectively. Left panel cells were categorized with Louvain clusters; the right panel cells were categorized by their source (SLE patient/healthy donors). **(B, F)** Bar plot of cell proportion in each B cell and monocyte subcluster, respectively. The dashed line represents the 75% cell proportion threshold. Both B cells (clusters 2, 6, 7) and monocytes (clusters 1, 7) have a unique cell subpopulation where SLE is predominant. **(C, G)** Heatmap of top five cluster-specific genes of each subclusters within B cells and monocytes, respectively. The color represents the expression level (blue: low, red: high). **(D, H)** UMAP plot of selected gene expression in re-clustered B cells and monocytes, respectively.

Detailed Ligand-Receptor Pair Alterations in SLE Patients

As the above results indicated that numerous signal patterns disappeared in SLE compared with healthy states, to find detailed information, we further explore the discrepancy of ligand–receptor pairs from all PBMC subpopulations (B cells, monocytes, T cells, natural killer cells, and dendritic cells) among healthy donor, adult SLE (aSLE), and child SLE (cSLE) groups (**Figures 5A–E**). We identified that eighty-seven ligand–receptor pairs disappeared in SLE patients, which were

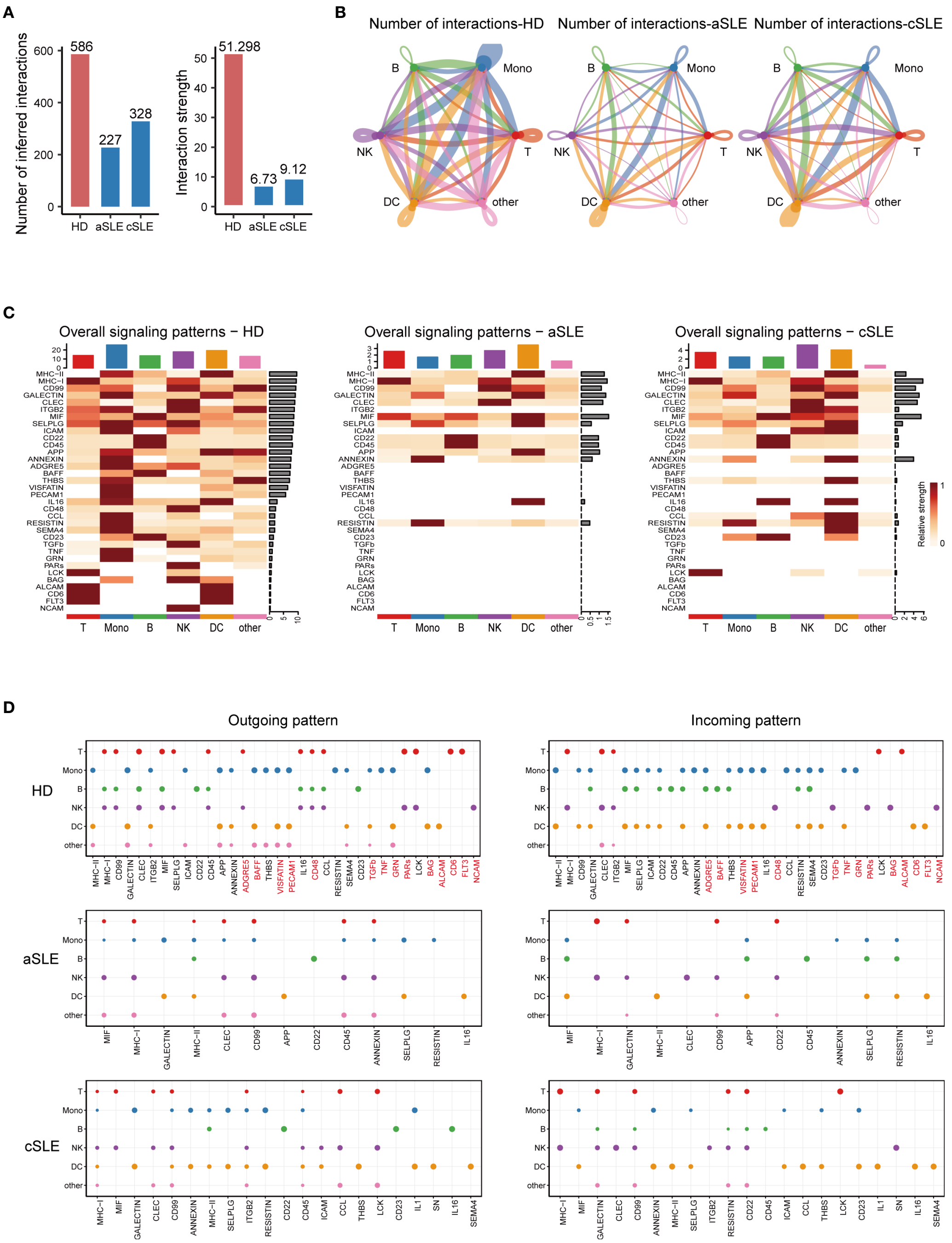


FIGURE 4 | CellChat analysis of whole PBMCs from SLE patients and healthy donors. **(A)** Bar plot of the overall difference among healthy donors (HD), adult SLE patients (aSLE), and child SLE patients (cSLE). The left panel shows the total number of interactions, and the right panel shows the interaction strength. **(B)** Circle plot of PBMC subpopulation among HD, aSLE, and cSLE. The line width: the connection strength; dark blue: monocytes, green: B cells, red: T cells, purple: natural killer cells, orange: dendritic cells and pink: other cells. These together revealed a weakened PBMC subpopulation cross talking and distinct signal pattern under SLE. **(C)** Heatmap reveals the overall signal pattern changes in the HD, aSLE, and cSLE groups, and the signal strength is scaled from white (no signal detected) to dark red (strong). **(D)** Dot plot for the emergence probability of signal outgoing (left panel) and incoming (right panel) patterns within each PBMC subpopulations among HD, aSLE, and cSLE. The dot size represents the p value. Patterns which specifically disappeared under disease state were marked with red. The total number of outgoing and incoming signal reduced significantly in SLE.

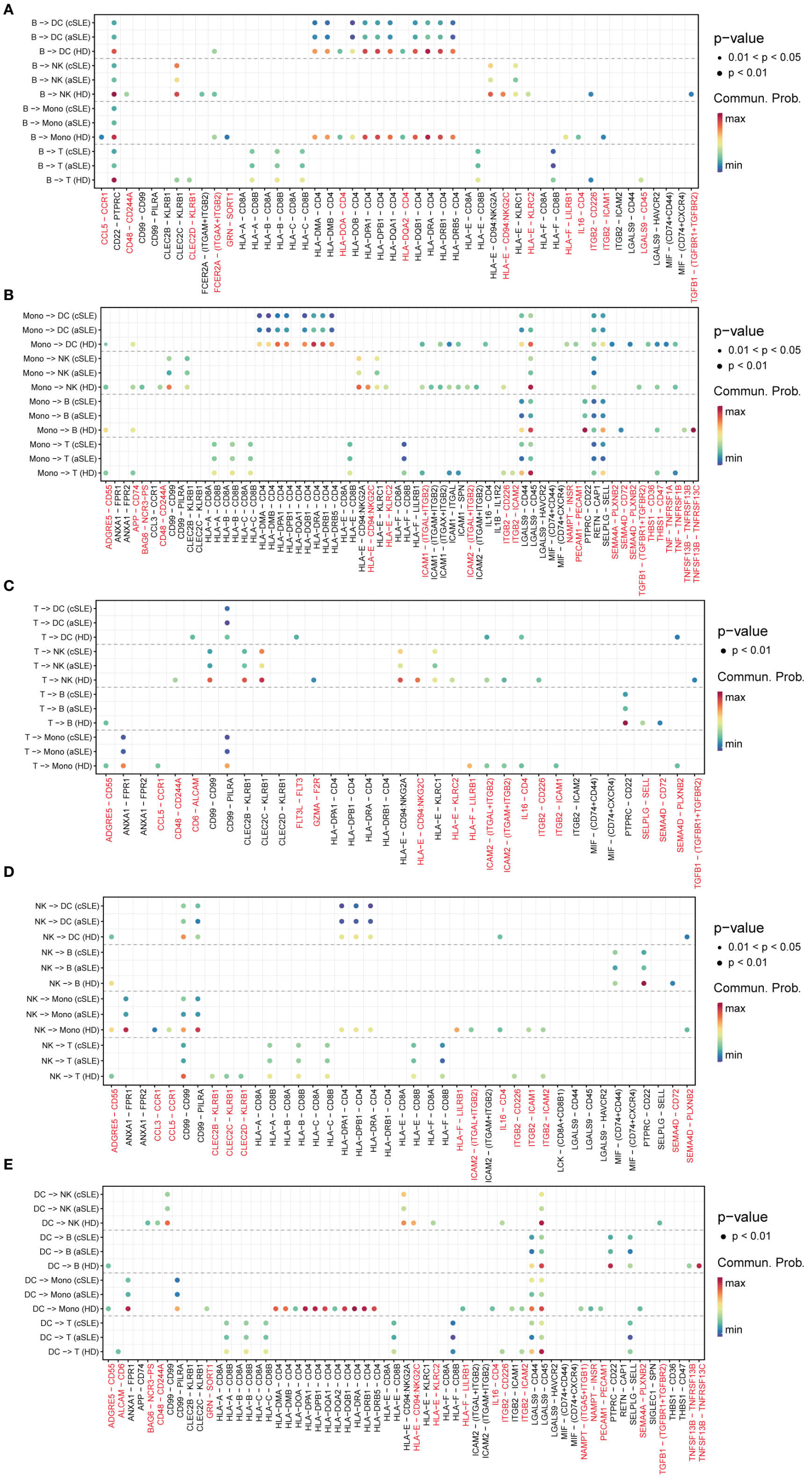


FIGURE 5 | Ligand–receptor pair alternation of SLE patients compared with healthy donors. Dot plot for the emergence probability of ligand–receptor pairs within each PBMC subpopulations **(A)** B cells, **(B)**monocytes, **(C)** T cells, **(D)** natural killer cells, **(E)** dendritic cells) among HD, aSLE, and cSLE. The dot color represents the probability. Pairs which specifically disappeared under disease state are marked with red.

composed of sixty-one genes. The frequency of each gene appeared at each PBMC subpopulation, as listed in **Table S2**. The genes which showed more than two kinds of PBMC subpopulation were recognized as significant ones to be selected as a second part of feature input for machine learning.

Among them, TGFBR1, TGFBR2, CCL5, CD48, CD244A, and CD72 have been reported to be closely related to the pathophysiologic processes of autoimmune diseases (41, 43, 45–47). For example, TGFBR1, TGFBR2, and CCL5 levels are negatively correlated with SLE development (43, 45). CD48, also known as SLAMF2, which could regulate both natural killer cells and cytotoxic CD8+ T cells (48), could protect mice from autoimmune nephritis (41), CD244A and CD72 were specifically decreased in monocytes and B cells during SLE development (47, 49). Interestingly, all these selected pairs are all in B cells or monocytes, suggesting the key roles of monocytes and B cells on the pathophysiologic processes of autoimmune diseases. All these findings were consistent with our results of integration analysis.

Efficient Machine Learning Models for Chronic Autoimmune Disease Diagnosis

To establish a mathematical model of unbiased random forest machine learning for SLE accurate diagnosis, sixty-seven top five cluster-specific genes derived from integration analysis and twenty-one significant genes identified *via* cellular cross-talking analysis were combined as feature input. The dataset GSE135779, containing 3.60 × 10^5 PBMCs (derived from 33 cSLE, 7 aSLE, and 11 healthy children, 5 healthy adults), was included to evaluate the diagnosis efficiency of our mathematical model.

The results indicated that our machine learning model could separate SLE and healthy status with acceptable accuracy (AUC = 0.776 ± 0.097, **Figure 6A**). The feature importance of our gene set for SLE is shown in **Figure 6C**. Considering the signal intensity of our gene sets and the denoising ability of machine learning, a further investigation was conducted to evaluate the disease distinguishing the efficiency of our mathematical model using bulk RNA-seq data. The bulk RNA-seq datasets (GSE72509, GSE183204), which include 99 SLE patients and 30 healthy donors were used in this investigation. The results indicated that our mathematical model has great adaptability (AUC = 0.998 ± 0.004, **Figure 6B**). The corresponding feature importance was also calculated (**Figure 6D**). This revealed that combined with the unbiased random forestry machine learning model, our gene sets rendered a powerful mathematical tool for distinguishing SLE.

It is reported that chronic autoimmune diseases including SLE and RA might share some similar cellular pathogeneses with MS (50). Thus, we investigated whether our machine learning model could efficiently distinguish RA and MS based on bulk RNA-seq data. Three datasets were included in this study, including a set of PBMC datasets (GSE90081, GSE183204) with 12 RA patients and 24 healthy donors, a synovial tissue dataset (GSE89408) with 152 RA patients and 28 healthy donors, and a PBMC dataset (GSE159225) with 20 relapse-and-remission MS patients, 10 secondary progressive MS patients, and 20 healthy donors.

Surprisingly, our machine learning model could separate patients with RA/MS and healthy donors with excellent accuracy in RA patients (AUC = 0.967 ± 0.099 in RA PBMC datasets, **Figure 7A**; AUC = 0.997 ± 0.006 in the RA synovial dataset, **Figure 7C**). For MS patients, our figure rendered an acceptable accuracy (AUC = 0.775 ± 0.236 in MS PBMC datasets, **Figure 7E**). The corresponding feature importance shown in **Figures 7B, D, F** illustrated that although our gene sets have extensive applicability and great accuracy for these diseases, each gene has different importance across each of these diseases. It suggested that our machine learning model requires a fine adjustment when applied to these diseases.

To determine the contribution of positive signals to the accuracy of our machine learning model, we obtain a public bulk RNA-seq dataset (GSE137143, 122 MS patients and 22 healthy donors), which consists of only CD14-positive monocytes. Unfortunately, the AUC value dropped to 0.673 ± 0.136, indicating that the accuracy sharply decreased (**Figure S4**). This result suggested that the distinguishing power of our model was reduced on account of a loss of positive signals, for example, the signals from B cells.

Diagnostic Accuracy Validation of the Machine Learning Model

To evaluate the diagnosis accuracy of our machine learning model, an independent cohort containing 120 SLE patients (GSE164457) and 41 healthy donors (derived from GSE169687) were enrolled into the study. The basic information and the gene expression pattern of objects within this cohort are shown in **Figures 8A, C**. Notably, the precision rate of our machine learning model diagnosis was 100% (120/120) and 92.7% (38/41) for SLE patients and healthy donors, respectively (**Figure 8B**). This result confirmed the diagnostic accuracy of our machine learning model, which suggested that it may be feasible to develop into an efficient tool for accurate disease diagnosis in the future.

DISCUSSION

We aimed to develop a feasible strategy for distinguishing patients with SLE and other major chronic autoimmune diseases in the early stage from healthy people. To achieve our purpose, the most crucial information that characterizes diseases should be filtered out first. From public single-cell RNA sequencing datasets, we found that B cells and monocytes were the only two subpopulations containing SLE-dominated clusters in the PBMCs of patients, which suggested that they might carry much stronger signals that indicate SLE than other PBMC subpopulations. To date, conclusions about the contribution of PBMC subpopulations to the development of SLE and other autoimmune diseases are not consistent, even when based on single-cell RNA sequencing data (51–55). Most studies mainly focus on specific disease aspects, which might result in imbalanced data selection, background noise interference, and biased conclusions. Hence, we selected the single-cell RNA sequencing data from over 1.50 × 10^5 cells for each category with a balanced ratio between patients and controls (approximately 1:1) to avoid rushing into any prejudicial conclusions.

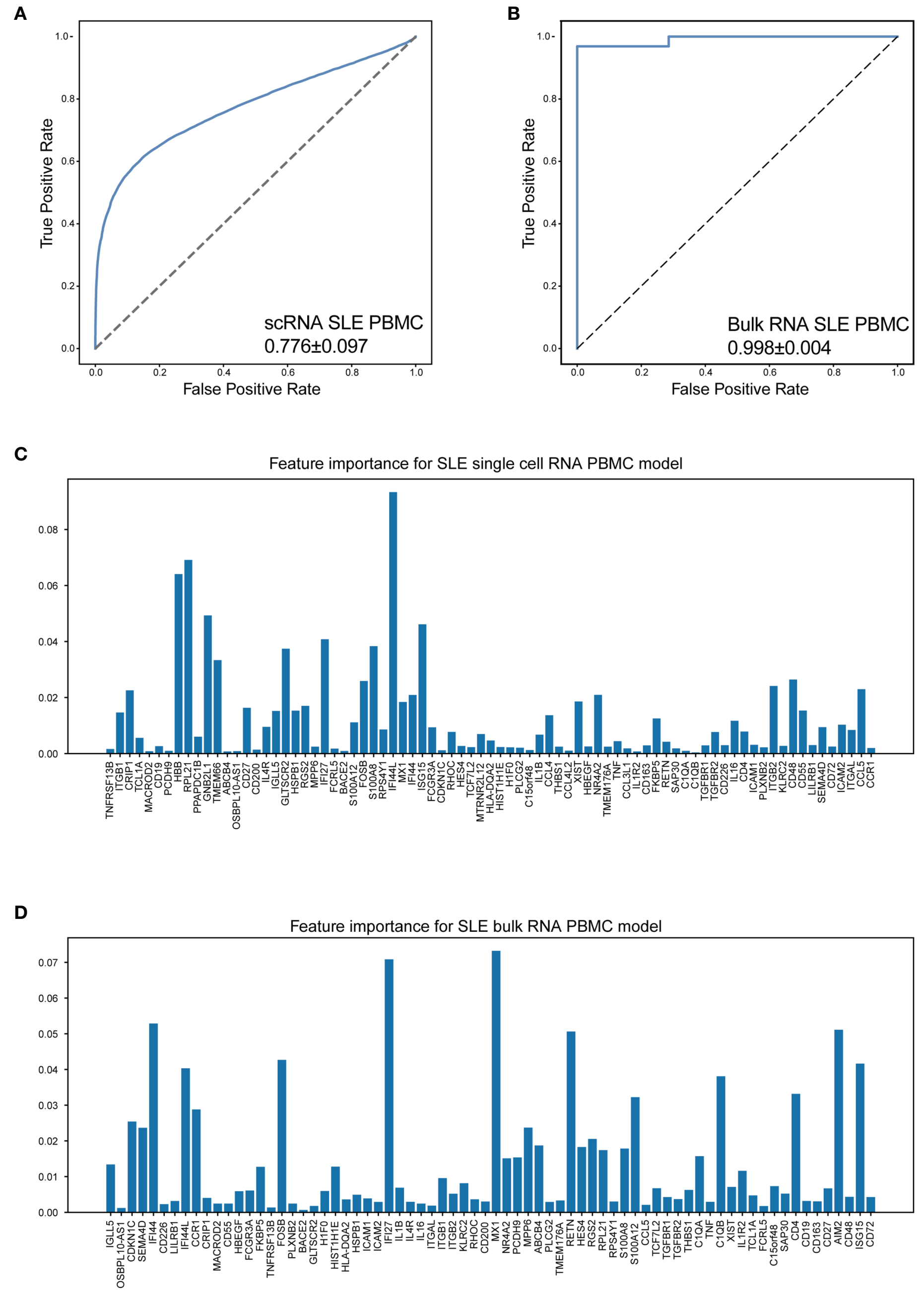


FIGURE 6 | Machine learning model accurately distinguish SLE. **(A)** The performance of distinguish SLE using scRNA-seq data of PBMCs (AUC = 0.776 ± 0.097). **(B)** The performance of distinguish SLE using bulk RNA-seq data of PBMCs (AUC = 0.998 ± 0.004). **(C, D)** Bar plot for the corresponding feature importance within the correlated model using scRNA-seq and bulk RNA-seq data, respectively. The bar length: feature importance.

Further investigation of differentially expressed genes revealed the details of the most significant information that marks a disease within B cells and monocytes. A few interferon-stimulated genes were active in the SLE-dominated B cells and monocytes, indicating that these cells might be a consequence of the inflammatory microenvironment. It is well known that the inflammatory microenvironment may be crucial to the progression of SLE and other chronic autoimmune diseases. Tsokos et al. reported that the production of autoantibodies triggered by both the innate and adaptive immune responses against self-antigens in SLE patients resulted in the accumulation of monocytes and activation of lymphocytes (56). Our results confirmed this suggestion. Interestingly, we found an activated naïve cluster of B cells in

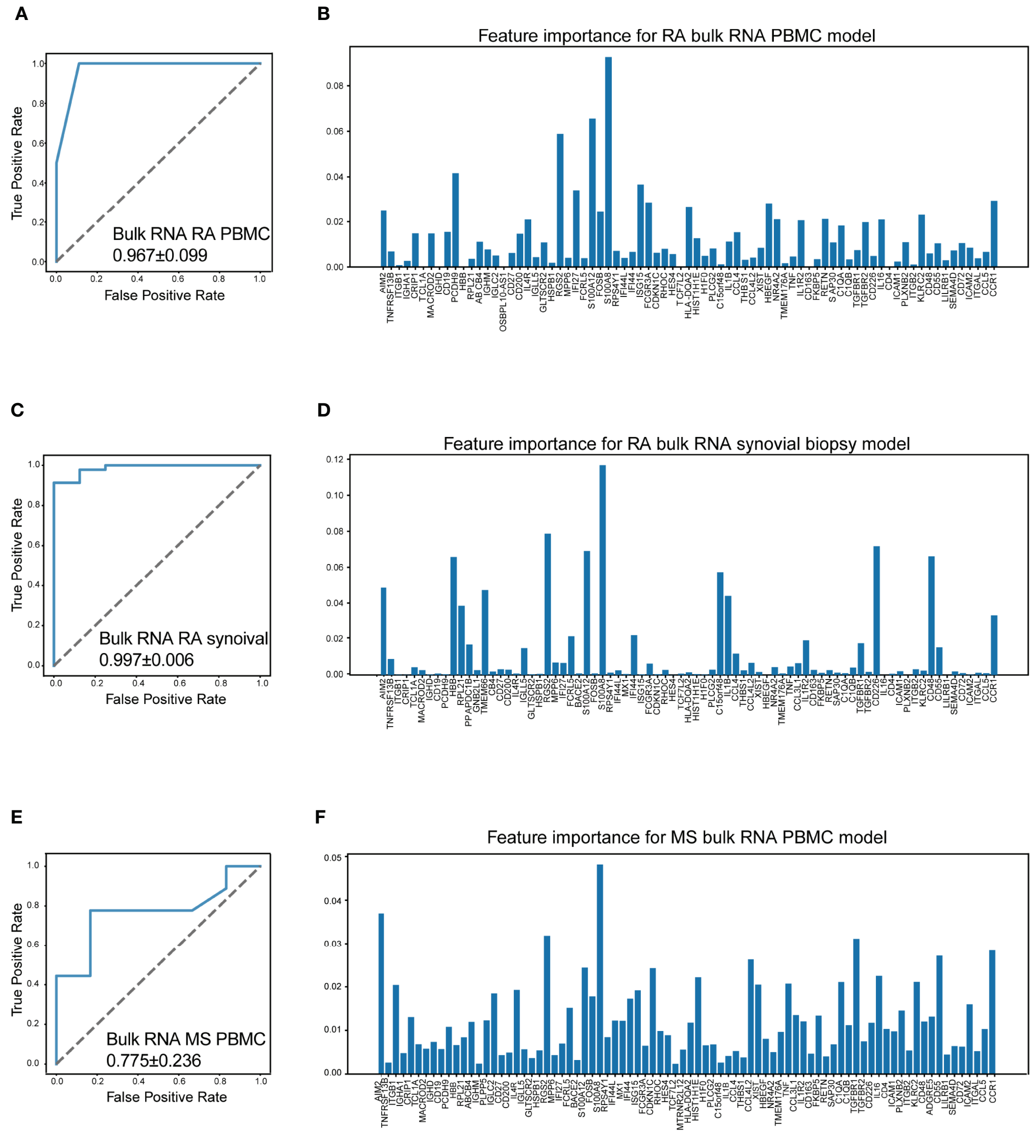
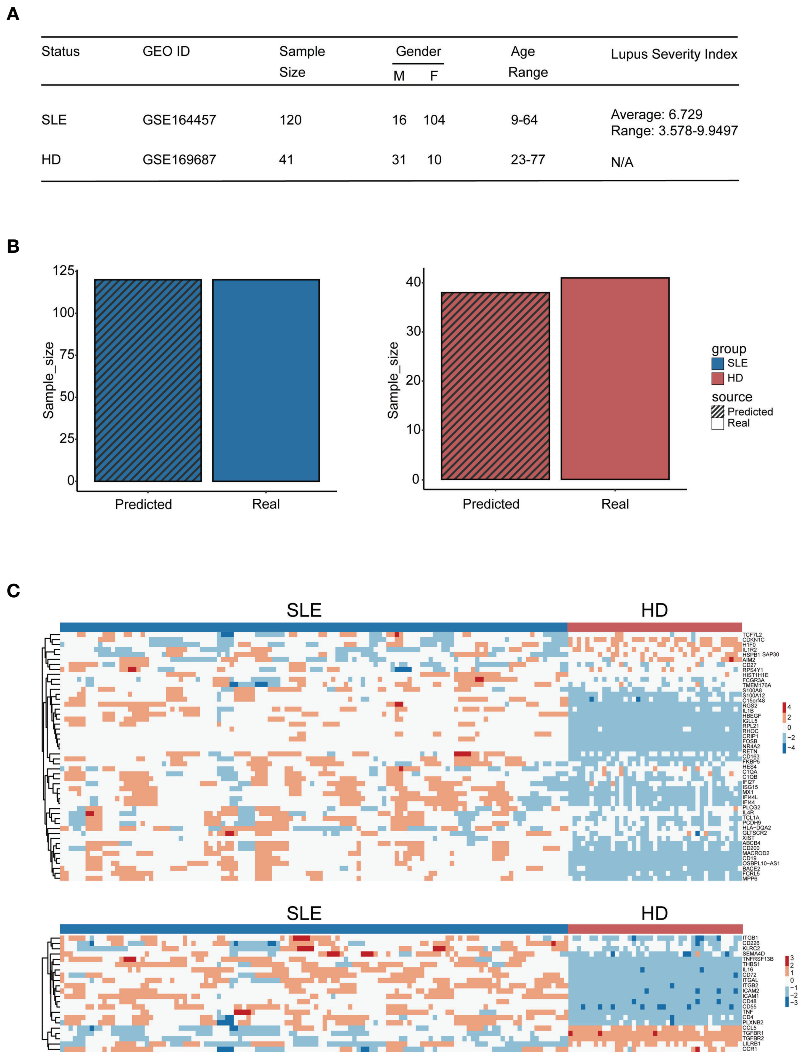


FIGURE 7 | Machine learning model accurately distinguish RA and MS. **(A)** The performance of distinguish RA (rheumatoid arthritis) using bulk RNA-seq data of PBMCs (AUC = 0.967 ± 0.099). **(B)** Bar plot for the feature importance with the correlated model. **(C)** The performance of distinguish RA using bulk RNA-seq data of synovial tissue (AUC = 0.997 ± 0.006). **(D)** Bar plot for the feature importance with the correlated model. **(E)** The performance of distinguish MS (multiple sclerosis) using bulk RNA-seq data of PBMCs (AUC = 0.775 ± 0.236). **(F)** Bar plot for the feature importance with the correlated model. The bar length: feature importance.

the SLE-dominated clusters. Recently, Jenks et al. reported a distinctive differentiation fate of autoreactive naïve B cells (39). This was similar to our finding and suggested that B cells should play an important role in the development of SLE.

All of the PBMC subpopulations were influenced mutually in the progression of chronic autoimmune diseases, and analyses based on individual subpopulations may lose important information of reciprocities that accounts for disease progression. Most current scRNA-seq data analysis tools focus on detailed categorizations and trajectories of cells (28, 57–59). Recently, bioinformatic tools (e.g., CellChat, CellPhoneDB, iTALK) were developed to infer cellular cross talking from scRNA-seq data, which make it possible to decipher reciprocities among cells under a single-cell level (57, 60–62). Therefore, we carried out cellular cross-talking analyses to reveal dynamic interactions across PBMC subpopulations and systematically decipher the etiology of diseases. Surprisingly, we found that the interactions among the PBMC subpopulations of SLE patients were weakened. It was reported that monocytes might contribute to the hyperactivity of B cells in SLE patients (63). A study also revealed that monocytes may function as a bridge during RA pathogenesis, and colocalization of CD14+ cells with CD4+ T effectors was found at sites of the inflamed rheumatoid synovium (64). Together, these reports illustrate that immune cells weave a



Status	GEO ID	Sample Size	Gender		Age Range	Lupus Severity Index
			M	F		
SLE	GSE164457	120	16	104	9-64	Average: 6.729 Range: 3.578-9.9497
HD	GSE169687	41	31	10	23-77	N/A

FIGURE 8 | Diagnostic accuracy validation of the machine learning model. **(A)** Table of cohort basic information. **(B)** Bar plot of the amount of SLE patients and healthy donors being distinguished accurately by the model (blue: SLE patients, red: HD); the bar with black stripe represents the model-predicted number, while the other represents the real number. **(C)** Heatmap of genes used for machine learning setup within the validation cohort (the upper panel: genes derived from the differential expression gene identification within integration analysis, the lower panel: genes derived from CellChat analysis).

network and that their interaction would provide significant information for autoimmune disease pathogenesis. Further detailed analysis revealed that the major changes occurred in B cells or monocytes, including FLT3, CD48, TNF, and TGF-beta signal patterns that have been reported to have a negative correlation with SLE development (41–44). Our results were consistent with previous studies on the variations in B cells (65–67) and monocytes (68–70) in SLE. Considering the repeatable results gained from our study, it should be convincing that the interactions among the PBMC subpopulations of SLE patients may be weakened, which could result in abnormal emergences or variations in signaling patterns within PBMCs.

Based on our finding of powerful information that characterizes diseases, we tried to establish a machine learning model to distinguish chronic autoimmune diseases. Several reports have proven that the random forest (RF) machine learning method would give a high accuracy in disease classification when abundant features were included (71, 72), and another reason for the random forest model was its interpretability—each gene contribution in the RF machine learning model was visible. Our area under curve (AUC) score for SLE indicates that our machine learning model has the potential to become an efficient tool for accurate diagnosis of SLE at the single-cell RNA level. Considering that the information we identified was not specific to the early stage of the disease, further optimization should be performed to identify the sensitive information in the early stage of the disease to strengthen the diagnostic power of our machine learning model.

Further investigation is also needed to evaluate the efficiency of our machine learning model using bulk RNA-sequencing data. Our AUC score illustrates that although other immune cell background noise might be introduced into RNA-seq data, the gene set still has high accuracy in distinguishing patients with the disease from healthy donors. This might be attributed to the low correlation between each gene since they were derived from the two different analysis frameworks, and this low gene correlation in turn increased the random forest model accuracy (73). Given the cost and convenience of bulk RNA sequencing, our results suggested that this machine learning model should be highly applicable going forward. In addition, our classification results for bulk RNA sequencing data of PBMCs and synovial tissues derived from RA and MS patients indicated that this machine learning model also showed high accuracy in distinguishing these diseases. Numerous studies have reported that chronic autoimmune diseases, such as SLE, RA, and MS, might share some similar cellular pathogeneses (46, 50, 74). Our findings further confirmed this viewpoint and suggested that this machine learning model with the information we filtered out might be powerful enough to discriminate patients with common chronic autoimmune diseases from healthy donors, not just SLE patients.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

Author contributions are shown as follows. Conception and design: AX, YM, and WH. Acquisition of data: XX, YQ. Analysis and interpretation of data: TW, LZ, JC and YM. Writing, review, and/or revision of the manuscript: all authors. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (81730005), the National Key Research and Development Program of China, Stem Cell and Translational Research (2018YFA0107203), the National Natural Science Foundation of China (32130046) and the National Natural Science Foundation of China (81970222) and the Key Scientific and Technological Projects of Guangdong Province (2016B0-90918040, 2017B020230004).

ACKNOWLEDGMENTS

The authors thanks Prof. Min Zhang of The First Affiliated Hospital, Sun Yat-sen University for manuscript revision.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.870531/full#supplementary-material

REFERENCES

1. Giacomelli R, Gorla R, Trotta F, Tirri R, Grassi W, Bazzichi L, et al. Quality of Life and Unmet Needs in Patients With Inflammatory Arthropathies: Results From the Multicentre, Observational Rapsodia Study. *Rheumatology* (2014) 54(5):792–7. doi: 10.1093/rheumatology/keu
2. Thamer M, Hernán MA, Zhang Y, Cotter D, Petri M. Prednisone, Lupus Activity, and Permanent Organ Damage. *J Rheumatol* (2009) 36(3):560–4. doi: 10.3899/jrheum.080828
3. Trapp BD, Peterson J, Ransohoff RM, Rudick R, Mörk S, Bö L. Axonal Transection in the Lesions of Multiple Sclerosis. *N Engl J Med* (1998) 338 (5):278–85. doi: 10.1056/nejm199801293380502
4. Batelaan NM, Bosman RC, Muntingh A, Scholten WD, Huijbregts KM, van Balkom A. Risk of Relapse After Antidepressant Discontinuation in Anxiety Disorders, Obsessive-Compulsive Disorder, and Post-Traumatic Stress Disorder: Systematic Review and Meta-Analysis of Relapse Prevention Trials. *BMJ (Clinical Res ed)* (2017) 358:j3927. doi: 10.1136/bmj.j3927
5. Kalincik T. Multiple Sclerosis Relapses: Epidemiology, Outcomes and Management. *A Systematic Review Neuroepidemiol* (2015) 44(4):199–214. doi: 10.1159/000382130
6. Arnaud L, Tektonidou MG. Long-Term Outcomes in Systemic Lupus Erythematosus: Trends Over Time and Major Contributors. *Rheumatol (Oxford England)* (2020) 59(Suppl5):v29–38. doi: 10.1093/rheumatology/keaa382
7. Doria A, Zen M, Canova M, Bettio S, Bassi N, Nalotto L, et al. Sle Diagnosis and Treatment: When Early Is Early. *Autoimmun Rev* (2010) 10(1):55–60. doi: 10.1016/j.autrev.2010.08.014
8. Sokka T, Kautiainen H, Pincus T, Verstappen SMM, Aggarwal A, Alten R, et al. Work Disability Remains a Major Problem in Rheumatoid Arthritis in the 2000s: Data From 32 Countries in the Quest-Ra Study. *Arthritis Res Ther* (2010) 12(2):R42. doi: 10.1186/ar2951
9. Cross M, Smith E, Hoy D, Carmona L, Wolfe F, Vos T, et al. The Global Burden of Rheumatoid Arthritis: Estimates From the Global Burden of Disease 2010 Study. *Ann rheumatic Dis* (2014) 73(7):1316–22. doi: 10.1136/annrheumdis-2013-204627
10. Kitas GD, Gabriel SE. Cardiovascular Disease in Rheumatoid Arthritis: State of the Art and Future Perspectives. *Ann rheumatic Dis* (2011) 70(1):8–14. doi: 10.1136/ard.2010.142133
11. Tamirou F, Arnaud L, Talarico R, Scirè CA, Alexander T, Amoura Z, et al. Systemic Lupus Erythematosus: State of the Art on Clinical Practice Guidelines. *RMD Open* (2019) 4(Suppl 1):e000793. doi: 10.1136/rmdopen-2018-000793
12. Solomon AJ, Corboy JR. The Tension Between Early Diagnosis and Misdiagnosis of Multiple Sclerosis. *Nat Rev Neurol* (2017) 13(9):567–72. doi: 10.1038/nrneurol.2017.106
13. De Cock D, Vanderschueren G, Meyfroidt S, Joly J, Westhovens R, Verschueren P. Two-Year Clinical and Radiologic Follow-Up of Early Ra Patients Treated With Initial Step Up Monotherapy or Initial Step Down Therapy With Glucocorticoids, Followed by a Tight Control Approach: Lessons From a Cohort Study in Daily Practice. *Clin Rheumatol* (2014) 33 (1):125–30. doi: 10.1007/s10067-013-2398-9
14. Mosca M, Costenbader KH, Johnson SR, Lorenzoni V, Sebastiani GD, Hoyer BF, et al. How Do Patients With Newly Diagnosed Systemic Lupus Erythematosus Present? A Multicenter Cohort of Early Systemic Lupus

Erythematosus to Inform the Development of New Classification Criteria. *Arthritis Rheumatol (Hoboken NJ)* (2019) 71(1):91–8. doi: 10.1002/art.40674

15. Brownlee WJ, Miller DH. Clinically Isolated Syndromes and the Relationship to Multiple Sclerosis. *J Clin Neurosci Off J Neurosurgical Soc Australasia* (2014) 21(12):2065–71. doi: 10.1016/j.jocn.2014.02.026
16. Solomon AJ, Naismith RT, Cross AH. Misdiagnosis of Multiple Sclerosis: Impact of the 2017 Mcdonald Criteria on Clinical Practice. *Neurology* (2019) 92(1):26–33. doi: 10.1212/wnl.0000000000006583
17. Piga M, Arnaud L. The Main Challenges in Systemic Lupus Erythematosus: Where Do We Stand? *J Clin Med* (2021) 10(2):243. doi: 10.3390/jcm10020243
18. Capecchi R, Puxeddu I, Pratesi F, Migliorini P. New Biomarkers in Sle: From Bench to Bedside. *Rheumatology* (2020) 59(Supplement_5):v12–v8. doi: 10.1093/rheumatology/keaa484
19. Rönnblom L, Leonard D. Interferon Pathway in Sle: One Key to Unlocking the Mystery of the Disease. *Lupus Sci Med* (2019) 6(1):e000270. doi: 10.1136/lupus-2018-000270
20. Mun S, Lee J, Park M, Shin J, Lim M-K, Kang H-G. Serum Biomarker Panel for the Diagnosis of Rheumatoid Arthritis. *Arthritis Res Ther* (2021) 23(1):31. doi: 10.1186/s13075-020-02405-7
21. Ziemssen T, Akgün K, Brück W. Molecular Biomarkers in Multiple Sclerosis. *J Neuroinflamm* (2019) 16(1):272. doi: 10.1186/s12974-019-1674-2
22. Touma Z, Gladman DD. Current and Future Therapies for Sle: Obstacles and Recommendations for the Development of Novel Treatments. *Lupus Sci Med* (2017) 4(1):e000239. doi: 10.1136/lupus-2017-000239
23. Pieterse E, van der Vlag J. Breaking Immunological Tolerance in Systemic Lupus Erythematosus. *Front Immunol* (2014) 5:164. doi: 10.3389/fimmu.2014.00164
24. Suurmond J, Diamond B. Autoantibodies in Systemic Autoimmune Diseases: Specificity and Pathogenicity. *J Clin Invest* (2015) 125(6):2194–202. doi: 10.1172/JCI78084
25. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive Integration of Single-Cell Data. *Cell* (2019) 177(7):1888–902.e21. doi: 10.1016/j.cell.2019.05.031
26. Wan S, Kim J, Won KJ. Sharp: Hyperfast and Accurate Processing of Single-Cell Rna-Seq Data *Via* Ensemble Random Projection. *Genome Res* (2020) 30 (2):205–13. doi: 10.1101/gr.254557.119
27. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan C-H, et al. Inference and Analysis of Cell-Cell Communication Using Cellchat. *Nat Commun* (2021) 12(1):1088. doi: 10.1038/s41467-021-21246-9
28. Nehar-Belaid D, Hong S, Marches R, Chen G, Bolisetty M, Baisch J, et al. Mapping Systemic Lupus Erythematosus Heterogeneity at the Single-Cell Level. *Nat Immunol* (2020) 21(9):1094–106. doi: 10.1038/s41590-020-0743-0
29. He B, Thomson M, Subramaniam M, Perez R, Ye CJ, Zou J. Cloudpred: Predicting Patient Phenotypes From Single-Cell Rna-Seq. *Pacific Symposium Biocomputing Pacific Symposium Biocomputing* (2022) 27:337–48. doi: 10.1142/9789811250477_0031
30. Luecken MD, Büttner M, Chaichoompu K, Danese A, Interlandi M, Mueller MF, et al. Benchmarking Atlas-Level Data Integration in Single-Cell Genomics. *Nat Methods* (2022) 19(1):41–50. doi: 10.1038/s41592-021-01336-8
31. Gholamy A, Kreinovich V, Kosheleva O. Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation. *J Intell Technol Appl* (2018) 2:105–11.
32. Becht E, McInnes L, Healy J, Dutertre C-A, Kwok IW, Ng LG, et al. Dimensionality Reduction for Visualizing Single-Cell Data Using Umap. *Nat Biotechnol* (2019) 37(1):38–44. doi: 10.1038/nbt.4314
33. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast Unfolding of Communities in Large Networks. *J Stat mechanics: Theory experiment* (2008) 2008(10):P10008. doi: 10.1088/1742-5468/2008/10/P10008
34. Hetherington-Rauth M, Bea JW, Blew RM, Funk JL, Hingle MD, Lee VR, et al. Relative Contributions of Lean and Fat Mass to Bone Strength in Young Hispanic and Non-Hispanic Girls. *Bone* (2018) 113:144–50. doi: 10.1016/j.bone.2018.05.023
35. Liu J, Berthier CC, Kahlenberg JM. Enhanced Inflammasome Activity in Systemic Lupus Erythematosus Is Mediated *Via* Type I Interferon-Induced Up-Regulation of Interferon Regulatory Factor 1. *Arthritis Rheumatol (Hoboken NJ)* (2017) 69(9):1840–9. doi: 10.1002/art.40166
36. Barbhaiya M, Liao KP. B-Cell Targeted Therapeutics in Systemic Lupus Erythematosus: From Paradox to Synergy? *Ann Internal Med* (2021) 174 (12):1747–8. doi: 10.7326/m21-4124
37. Chiche L, Jourde-Chiche N, Whalen E, Presnell S, Gersuk V, Dang K, et al. Modular Transcriptional Repertoire Analyses of Adults With Systemic Lupus Erythematosus Reveal Distinct Type I and Type Ii Interferon Signatures. *Arthritis Rheumatol (Hoboken NJ)* (2014) 66(6):1583–95. doi: 10.1002/art.38628
38. Sanz I, Wei C, Jenks SA, Cashman KS, Tipton C, Woodruff MC, et al. Challenges and Opportunities for Consistent Classification of Human B Cell and Plasma Cell Populations. *Front Immunol* (2019) 10:2458. doi: 10.3389/fimmu.2019.02458
39. Jenks SA, Cashman KS, Zumaquero E, Marigorta UM, Patel AV, Wang X, et al. Distinct Effector B Cells Induced by Unregulated Toll-Like Receptor 7 Contribute to Pathogenic Responses in Systemic Lupus Erythematosus. *Immunity* (2018) 49(4):725–39.e6. doi: 10.1016/j.immuni.2018.08.015
40. Zannas AS, Jia M, Hafner K, Baumert J, Wiechmann T, Pape JC, et al. Epigenetic Upregulation of Fkbp5 by Aging and Stress Contributes to Nf-κb–Driven Inflammation and Cardiovascular Risk. *Proc Natl Acad Sci* (2019) 116 (23):11370. doi: 10.1073/pnas.1816847116
41. Koh AE, Njoroge SW, Feliu M, Cook A, Selig MK, Latchman YE, et al. The Slam Family Member Cd48 (Slamf2) Protects Lupus-Prone Mice From Autoimmune Nephritis. *J Autoimmun* (2011) 37(1):48–57. doi: 10.1016/j.jaut.2011.03.004
42. Aringer M, Smolen JS. The Role of Tumor Necrosis Factor-Alpha in Systemic Lupus Erythematosus. *Arthritis Res Ther* (2008) 10(1):202. doi: 10.1186/ar2341
43. Rekik R, Smiti Khanfir M, Larbi T, Zamali I, Beldi-Ferchiou A, Kammoun O, et al. Impaired Tgf-β Signaling in Patients With Active Systemic Lupus Erythematosus Is Associated With an Overexpression of Il-22. *Cytokine* (2018) 108:182–9. doi: 10.1016/j.cyto.2018.04.011
44. Yuan X, Qin X, Wang D, Zhang Z, Tang X, Gao X, et al. Mesenchymal Stem Cell Therapy Induces Flt3l and Cd1c+ Dendritic Cells in Systemic Lupus Erythematosus Patients. *Nat Commun* (2019) 10(1):2498. doi: 10.1038/s41467-019-10491-8
45. Zhu H, Mi W, Luo H, Chen T, Liu S, Raman I, et al. Whole-Genome Transcription and DNA Methylation Analysis of Peripheral Blood Mononuclear Cells Identified Aberrant Gene Regulation Pathways in Systemic Lupus Erythematosus. *Arthritis Res Ther* (2016) 18(1):162. doi: 10.1186/s13075-016-1050-x
46. Ma W-T, Gao F, Gu K, Chen D-K. The Role of Monocytes and Macrophages in Autoimmune Diseases: A Comprehensive Review. *Front Immunol* (2019) 10:1140. doi: 10.3389/fimmu.2019.01140
47. Tsubata T. Cd72 Is a Negative Regulator of B Cell Responses to Nuclear Lupus Self-Antigens and Development of Systemic Lupus Erythematosus. *Immune Netw* (2019) 19(1):e1–e. doi: 10.4110/in.2019.19.e1
48. Kis-Toth K, Comte D, Karampetsou MP, Kyttaris VC, Kannan L, Terhorst C, et al. Selective Loss of Signaling Lymphocytic Activation Molecule Family Member 4-Positive Cd8+ T Cells Contributes to the Decreased Cytotoxic Cell Activity in Systemic Lupus Erythematosus. *Arthritis Rheumatol (Hoboken NJ)* (2016) 68(1):164–73. doi: 10.1002/art.39410
49. Mak A, Thornhill SI, Lee HY, Lee B, Poidinger M, Connolly JE, et al. Brief Report: Decreased Expression of Cd244 (Slamf4) on Monocytes and Platelets in Patients With Systemic Lupus Erythematosus. *Clin Rheumatol* (2018) 37 (3):811–6. doi: 10.1007/s10067-017-3698-2
50. Lee DSW, Rojas OL, Gommerman JL. B Cell Depletion Therapies in Autoimmune Disease: Advances and Mechanistic Insights. *Nat Rev Drug Discovery* (2021) 20(3):179–99. doi: 10.1038/s41573-020-00092-2
51. Trzupek D, Lee M, Hamey F, Wicker LS, Todd JA, Ferreira RC. Single-Cell Multi-Omics Analysis Reveals Ifn-Driven Alterations in T Lymphocytes and Natural Killer Cells in Systemic Lupus Erythematosus. *Wellcome Open Research* (2021) 6(149):149. doi: 10.12688/wellcomeopenres.16883.1
52. Dutertre CA, Becht E, Irac SE, Khalilnezhad A, Narang V, Khalilnezhad S, et al. Single-Cell Analysis of Human Mononuclear Phagocytes Reveals Subset-Defining Markers and Identifies Circulating Inflammatory Dendritic Cells. *Immunity* (2019) 51(3):573–89.e8. doi: 10.1016/j.immuni.2019.08.008
53. McHugh J. Newly Defined Pro-Inflammatory Dc Subset Expanded in Sle. *Nat Rev Rheumatol* (2019) 15(11):637–. doi: 10.1038/s41584-019-0311-x
54. Nakano M, Iwasaki Y, Fujio K. Transcriptomic Studies of Systemic Lupus Erythematosus. *Inflammation Regeneration* (2021) 41(1):11. doi: 10.1186/s41232-021-00161-y

55. Kondo Y, Yokosawa M, Kaneko S, Furuyama K, Segawa S, Tsuboi H, et al. Review: Transcriptional Regulation of Cd4+ T Cell Differentiation in Experimentally Induced Arthritis and Rheumatoid Arthritis. *Arthritis Rheumatol (Hoboken NJ)* (2018) 70(5):653–61. doi: 10.1002/art.40398
56. Tsokos GC, Lo MS, Costa Reis P, Sullivan KE. New Insights Into the Immunopathogenesis of Systemic Lupus Erythematosus. *Nat Rev Rheumatol* (2016) 12(12):716–30. doi: 10.1038/nrrheum.2016.186
57. Jin W, Yang Q, Peng Y, Yan C, Li Y, Luo Z, et al. Single-Cell Rna-Seq Reveals Transcriptional Heterogeneity and Immune Subtypes Associated With Disease Activity in Human Myasthenia Gravis. *Cell Discov* (2021) 7(1):85. doi: 10.1038/s41421-021-00314-w
58. Heng JS, Hackett SF, Stein-O'Brien GL, Winer BL, Williams J, Goff LA, et al. Comprehensive Analysis of a Mouse Model of Spontaneous Uveoretinitis Using Single-Cell Rna Sequencing. *Proc Natl Acad Sci* (2019) 116(52):26734. doi: 10.1073/pnas.1915571116
59. Zakharov PN, Hu H, Wan X, Unanue ER. Single-Cell Rna Sequencing of Murine Islets Shows High Cellular Complexity at All Stages of Autoimmune Diabetes. *J Exp Med* (2020) 217(6):e20192362. doi: 10.1084/jem.20192362
60. Li H, Gao Y, Xie L, Wang R, Duan R, Li Z, et al. Prednisone Reprograms the Transcriptional Immune Cell Landscape in Cns Autoimmune Disease. *Front Immunol* (2021) 12:739605. doi: 10.3389/fimmu.2021.739605
61. Li T, Shen K, Li J, Leung SWS, Zhu T, Shi Y. Glomerular Endothelial Cells Are the Coordinator in the Development of Diabetic Nephropathy. *Front Med* (2021) 8:655639. doi: 10.3389/fmed.2021.655639
62. Stephenson E, Reynolds G, Botting RA, Calero-Nieto FJ, Morgan MD, Tuong ZK, et al. Single-Cell Multi-Omics Analysis of the Immune Response in Covid-19. *Nat Med* (2021) 27(5):904–16. doi: 10.1038/s41591-021-01329-2
63. Blanco P, Palucka AK, Gill M, Pascual V, Banchereau J. Induction of Dendritic Cell Differentiation by Ifn-Alpha in Systemic Lupus Erythematosus. *Sci (New York NY)* (2001) 294(5546):1540–3. doi: 10.1126/science.1064890
64. Fonseca JE, Edwards JC, Blades S, Goulding NJ. Macrophage Subpopulations in Rheumatoid Synovium: Reduced Cd163 Expression in Cd4+ T Lymphocyte-Rich Microenvironments. *Arthritis Rheumatism* (2002) 46 (5):1210–6. doi: 10.1002/art.10207
65. Faridi MH, Khan SQ, Zhao W, Lee HW, Altintas MM, Zhang K, et al. Cd11b Activation Suppresses Tlr-Dependent Inflammation and Autoimmunity in Systemic Lupus Erythematosus. *J Clin Invest* (2017) 127(4):1271–83. doi: 10.1172/jci88442
66. Haynes WA, Haddon DJ, Diep VK, Khatri A, Bongen E, Yiu G, et al. Integrated, Multicohort Analysis Reveals Unified Signature of Systemic Lupus Erythematosus. *JCI Insight* (2020) 5(4):e122312. doi: 10.1172/jci.insight.122312
67. Maeda N, Sekigawa I, Iida N, Matsumoto M, Hashimoto H, Hirose S. Relationship Between Cd4+/Cd8+ T Cell Ratio and T Cell Activation in Systemic Lupus Erythematosus. *Scandinavian J Rheumatol* (1999) 28(3):166–70. doi: 10.1080/03009749950154248
68. Park JK, Lee YJ, Park JS, Lee EB, Song YW. Cd47 Potentiates Inflammatory Response in Systemic Lupus Erythematosus. *Cells* (2021) 10(5):1151. doi: 10.3390/cells10051151
69. Sabry A, Sheashaa H, El-Husseini A, El-Dahshan K, Abdel-Rahim M, Elbasyouni SR. Intercellular Adhesion Molecules in Systemic Lupus Erythematosus Patients With Lupus Nephritis. *Clin Rheumatol* (2007) 26 (11):1819–23. doi: 10.1007/s10067-007-0580-7
70. Rullo OJ, Tsao BP. Recent Insights Into the Genetic Basis of Systemic Lupus Erythematosus. *Ann rheumatic Dis* (2013) 72 Suppl 2(0 2):ii56–61. doi: 10.1136/annrheumdis-2012-202351
71. Cao Y, Wang L, Ke S, Villafuerte Gálvez JA, Pollock NR, Barrett C, et al. Fecal Mycobiota Combined With Host Immune Factors Distinguish Clostridioides Difficile Infection From Asymptomatic Carriage. *Gastroenterology* (2021) 160 (7):2328–39.e6. doi: 10.1053/j.gastro.2021.02.069
72. Xiang J, Shi M, Fiala MA, Gao F, Rettig MP, Uy GL, et al. Machine Learning-Based Scoring Models to Predict Hematopoietic Stem Cell Mobilization in Allogeneic Donors. *Blood Adv* (2021) 6(7):1991–2000. doi: 10.1182/bloodadvances.2021005149
73. Fawagreh K, Gaber MM, Elyan E. Random Forests: From Early Developments to Recent Advancements. *Syst Sci Control Eng* (2014) 2(1):602–9. doi: 10.1080/21642583.2014.956265
74. Hirose S, Lin Q, Ohtsuji M, Nishimura H, Verbeek JS. Monocyte Subsets Involved in the Development of Systemic Lupus Erythematosus and Rheumatoid Arthritis. *Int Immunol* (2019) 31(11):687–96. doi: 10.1093/intimm/dxz036

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Immunology

MINI REVIEW
published: 28 April 2022
doi: 10.3389/fimmu.2022.845655

Roles Played by Stress-Induced Pathways in Driving Ethnic Heterogeneity for Inflammatory Skin Diseases

Taylor A. Jamerson[1], Qinmengge Li[2], Sutharzan Sreeskandarajan[1], Irina V. Budunova[3,4], Zhi He[2], Jian Kang[2], Johann E. Gudjonsson[1], Matthew T. Patrick[1] and Lam C. Tsoi[1,2,5]*

[1] Department of Dermatology, Michigan Medicine, University of Michigan, Ann Arbor, MI, United States, [2] Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI, United States, [3] Department of Dermatology, Northwestern Medicine, Northwestern University, Chicago, IL, United States, [4] Department of Urology, Northwestern Medicine, Northwestern University, Chicago, IL, United States, [5] Department of Computational Medicine and Bioinformatics, Michigan Medicine, University of Michigan, Ann Arbor, MI, United States

OPEN ACCESS

Edited by:
Lisa G. Rider,
National Institute of Environmental Health Sciences (NIH), United States

Reviewed by:
Lennart Matthias Roesner,
Hannover Medical School, Germany

***Correspondence:**
Lam C. Tsoi
alextsoi@med.umich.edu

Specialty section:
This article was submitted to Autoimmune and Autoinflammatory Disorders, a section of the journal Frontiers in Immunology

Received: 30 December 2021
Accepted: 23 March 2022
Published: 28 April 2022

Citation:
Jamerson TA, Li Q, Sreeskandarajan S, Budunova IV, He Z, Kang J, Gudjonsson JE, Patrick MT and Tsoi LC (2022) Roles Played by Stress-Induced Pathways in Driving Ethnic Heterogeneity for Inflammatory Skin Diseases. Front. Immunol. 13:845655. doi: 10.3389/fimmu.2022.845655

Immune-mediated skin conditions (IMSCs) are a diverse group of autoimmune diseases associated with significant disease burden. Atopic dermatitis and psoriasis are among the most common IMSCs in the United States and have disproportionate impact on racial and ethnic minorities. African American patients are more likely to develop atopic dermatitis compared to their European American counterparts; and despite lower prevalence of psoriasis among this group, African American patients can suffer from more extensive disease involvement, significant post-inflammatory changes, and a decreased quality of life. While recent studies have been focused on understanding the heterogeneity underlying disease mechanisms and genetic factors at play, little emphasis has been put on the effect of psychosocial or psychological stress on immune pathways, and how these factors contribute to differences in clinical severity, prevalence, and treatment response across ethnic groups. In this review, we explore the heterogeneity of atopic dermatitis and psoriasis between African American and European American patients by summarizing epidemiological studies, addressing potential molecular and environmental factors, with a focus on the intersection between stress and inflammatory pathways.

Keywords: African America, psoriasis, atopic dermatitis, stress, minority

INTRODUCTION

Over 84 million Americans are impacted by at least one skin disease (1), posing significant health and economic burden. Atopic dermatitis (AD) and psoriasis are among the most common and widely studied immune-mediated skin conditions (IMSCs). With recent advances in genomic technology, studies over the past decade (2–8) have focused on elucidating the underlying mechanisms of disease pathogenesis for AD and psoriasis. More specifically, these studies have explored the genetic and molecular factors associated with the disease pathophysiology. However, the primary racial makeup of these studies has been predominantly European American (EA).

While AD and psoriasis affect populations of different origins, their burden is exacerbated in some ethnic minority groups (9–11). Yet, there is a paucity of molecular studies describing the driving factors behind ethnic heterogeneity in the severity, presentation, and predominance of IMSCs.

Environmental factors, such as stress, are known to play roles in shaping the onset and clinical severity of IMSCs (12–14). While chronic stressors may impact any individual, unique psychosocial factors such as racism, discrimination, and acculturative stress (anxiety or tension related to efforts to adapt to the values of dominant culture within a society) are unique among ethnic minorities (15). Studies from the US Department of Health and Human Services Office of Minority Health show that African American (AA) adults living below the poverty line are twice as likely to experience psychological distress (16) and the downstream effects of these stressors can result in decreased likelihood of receiving medical care (17, 18). It is therefore important to understand the influence of stressors, including psychosocial stress, on the pathogenesis of IMSCs, especially as this may play role in the ethnic differences seen across common IMSCs.

In this review, we explore the heterogeneity in AD and psoriasis across AA and EA patients by summarizing epidemiological studies, as well as the potential molecular and environmental factors involved in disease pathogenesis. We also place particular focus on the intersections between known stress pathways and IMSC inflammatory pathways in the literature.

ATOPIC DERMATITIS

Epidemiology

Atopic dermatitis (AD) is an inflammatory skin condition affecting more than 18 million adults in the United States (19, 20). This condition classically presents with pruritic, erythematous plaques involving the flexor surfaces, particularly in the antecubital and popliteal fossae. In Fitzpatrick skin types IV through VI however, eczematous patches often appear brown, purple, or ashen grey in color (11). Clinically, AD often presents with greater involvement of the flexural surfaces in adults, however patients of African descent are more likely to present with more prominent involvement of the extensor surfaces (21). Previous epidemiological studies using self-reported ethnic information highlight a slightly higher prevalence of AD in AA patients when compared with EAs (19.3% versus 16.1%) (22). This predominance is also seen at young age, with AA children found to be 1.7 times more likely to develop AD compared to their EA counterparts, even after adjusting for health insurance and socioeconomic status (11). In addition to reported racial differences in AD prevalence, AA children as a group has been reported to have more severe disease than EA children (23); the study also suggested structural racism or the increased proportion of AA children living in lower income, segregated communities with exposure to greater environmental stressors, are associated with disease severity. To further validate and understand the epidemiological factors involved, we studied the demographical variables from an insurance-claim database, Optum Clinformatics Data Mart (CDM), consisting of 1,458,417 AD patient records across 2014 to 2018 (**Table 1**). As expected, the association of AD-related clinical visits was significantly stronger at younger age (<18 years) for all ethnic groups compared with our reference age group (18-65 years) (OR:1.60, 1.95, 2.38, 1.92 for EA, Hispanic, Asian and AA populations, respectively). The older patient population group (>65 years) also had significantly stronger association with AD clinical visits (OR=2.18, 1.76, 1.52, 1.75 for the same four ethnic groups, respectively). The data importantly highlights that gender factors had the largest effect sizes in AA (e.g. OR=1.43 for female) compared to the other ethnic groups (OR between 1.23-1.35 for female). Patients with higher income are also associated with higher prevalance of clinical visit for AD.

Genetics

While the cause of AD is complex, studies over the last decade have provided insights into genetic and environmental factors associated with disease pathogenesis and severity. Filaggrin (*FLG*), a protein encoding gene from the epidermal differentiation complex (EDC) and expressed in the keratinized layer of the epidermis, plays an important role in skin barrier function, for example by promoting keratinocyte differentiation and rapid cell death (24). FLG expression is immune-modulated by both aryl hydrocarbon receptor signaling and cytokines, resulting in dysregulation within the lesional skin (25). The locus harboring *FLG* has been identified as one of the strongest genetic signals associated with AD (26). Carriers of the *FLG* loss-of-function (LOF) variants have increased odds (3-fold) of having AD. While LOF *FLG* mutations are risk factors for the development of AD in patients of European and Asian descent, this association has not been reported among individuals of African descent. In fact, loss of function *FLG* mutations are thought to be less common in AA patients with AD when compared to EAs (27), and a recent study suggests that the common *FLG* mutation found in AA patients are distinct from those in EA and Asians (28). Nevertheless, decreased expression and mutations of filaggrin-2 (*FLG2*), also from EDC, have been associated with persistent symptoms of AD in AA, while such variations are absent or infrequently found in AD patients of European ancestry (29). The first genome-wide significant association at rs3811419 for AD in AA patients was found to be an expression quantitative trait loci (eQTL) for *THEM4* (a gene associated with allergy), and *FLG-AS1* (a non-coding RNA that overlaps the filaggrin gene) in blood. The association between the risk allele of this eQTL locus and increased expression of *FLG-AS1* can potentially offer an alternative mechanism explaining skin barrier deficiency in AA, although further research is required (30).

While AA patients are more likely to present with more severe AD than EAs patients (22), a recent study challenged the perception that racial heterogeneity in the severity of AD is genetically driven (31). This study found that observed differences of AD in AA patients, including disease severity, were not associated with a continuous measure of African genetic ancestry. This finding supports epidemiologic, rather than genetic, associations with disease severity and suggests the

TABLE 1 | Risk factors for atopic dermatitis and psoriasis stratified by different ethnic groups.

		Atopic Dermatitis				Psoriasis			
		Odds Ratio	95% CI Lower bound	95% CI Upper bound	Significance	Odds Ratio	95% CI Lower bound	95% CI Upper bound	Significance
Obesity	**EA**	1.49	1.48	1.51	***	1.89	1.86	1.91	***
	Hispanic	1.47	1.44	1.51	***	1.91	1.84	1.97	***
	Asian	1.32	1.25	1.40	***	1.98	1.83	2.14	***
	AA	1.55	1.51	1.59	***	1.75	1.68	1.81	***
Age (18-65 as ref)	**EA (y)**	1.60	1.58	1.62	***	0.18	0.17	0.18	***
	EA (o)	2.18	2.16	2.21	***	1.80	1.77	1.82	***
	Hispanic (y)	1.95	1.88	2.03	***	0.20	0.18	0.22	***
	Hispanic (o)	1.76	1.69	1.84	***	1.45	1.37	1.54	***
	Asian (y)	2.38	2.31	2.46	***	0.22	0.19	0.24	***
	Asian (o)	1.52	1.48	1.56	***	2.17	2.10	2.26	***
	AA (y)	1.92	1.87	1.97	***	0.21	0.20	0.23	***
	AA (o)	1.75	1.70	1.79	***	1.57	1.52	1.63	***
Gender (M as ref)	**EA**	1.35	1.34	1.36	***	1.10	1.09	1.11	***
	Hispanic	1.34	1.31	1.36	***	0.97	0.95	1.00	.
	Asian	1.23	1.19	1.27	***	0.84	0.80	0.89	***
	AA	1.43	1.40	1.46	***	1.14	1.10	1.18	***
$40K-$49K	**EA**	1.08	1.06	1.10	***	1.02	0.99	1.05	.
	Hispanic	1.05	1.01	1.09	**	1.00	0.94	1.06	.
	Asian	1.03	0.95	1.12	.	1.11	0.97	1.27	.
	AA	1.11	1.07	1.15	***	0.98	0.92	1.04	.
$50K-$59K	**EA**	1.13	1.11	1.15	***	1.07	1.05	1.10	***
	Hispanic	1.11	1.07	1.15	***	1.07	1.01	1.14	*
	Asian	1.09	1.01	1.18	*	1.14	1.01	1.30	*
	AA	1.22	1.18	1.27	***	1.00	0.94	1.06	
$60K-$74K	**EA**	1.21	1.19	1.23	***	1.14	1.11	1.16	***
	Hispanic	1.18	1.14	1.22	***	1.15	1.09	1.21	***
	Asian	1.09	1.02	1.17	**	1.13	1.01	1.25	*
	AA	1.32	1.27	1.37	***	1.11	1.05	1.18	***
$75K-$99K	**EA**	1.31	1.29	1.33	***	1.23	1.20	1.25	***
	Hispanic	1.31	1.27	1.35	***	1.25	1.19	1.31	***
	Asian	1.13	1.06	1.20	***	1.24	1.13	1.37	***
	AA	1.37	1.32	1.42	***	1.26	1.19	1.33	***
$100K+	**EA**	1.71	1.69	1.73	***	1.52	1.49	1.55	***
	Hispanic	1.61	1.57	1.65	***	1.62	1.55	1.69	***
	Asian	1.25	1.19	1.32	***	1.37	1.26	1.49	***
	AA	1.63	1.58	1.68	***	1.56	1.48	1.65	***

*The demographical variables from an insurance-claim database (CDM) of 1,458,417 atopic dermatitis and 272,913 psoriatic patient records were analyzed across 2014 to 2018. For age, individuals of 18-65 years were used as reference for <18 years age (y) and >65 years age (o) groups; for socio-economic status, individuals with household income <$40,000 were used as reference. For significance level: *0.01<p ≤ 0.05, **0.001<p ≤ 0.01, ***p ≤ 0.001. Income is reported in the United States dollar (USD).*

Optum Clinformatics Data Mart (CDM) is a claim-based Electronic Health Record (EHR) database containing demographic, diagnosis, pharmacy, and lab analyte records for > 63 million de-identified patients from 2001 to 2018 in US (https://www.optum.com/business/solutions/life-sciences/real-world-data/claims-data.html). We restricted our study to only consider recent patient visit between 2014-2018, and implemented a case-control study framework to further decrease the size of patients without the target disease. In the analysis, we used all the AD or psoriasis patients as our case sample, and randomly drew 3 million patients from the rest of the data and included only the patients that visited and filed claims between 2014-2018. We then fitted logistic regressions on the AD or psoriasis indicator adjusting for obesity, age, gender, household income and race. Obesity, age, gender, household income covariates are coded by the reference cell coding scheme, with reference levels to be non-obesity, 18-65, male and below $40,000 respectively.

potential roles of other factors such as environmental components (e.g., stress, pollution, social determinants of health) in disease heterogeneity and their influence on disease pathogenesis.

Stress and Immunologic Parameters

Though most studies describing the molecular signature of AD have been conducted in patients of European ancestry, a recent study by Wongvibulsin et al. confirmed previously reported Th_2/Th_{22} skewing in AA patients with AD, along with upregulated Th_1 cytokines in lesional skin of AD, contribute to the increased disease severity in AA patients (32). The authors also reported elevated serum C-reactive protein (CRP), ferritin, and blood eosinophils in AA patients when compared to EA patients with AD. In addition, Schmeer et al. measured serum CRP levels across ethnic groups in children aged 2 to 10 years and found significantly higher serum measurements in AA and Hispanic children when compared to EA children (33). In adults, increased activity of two essential pro-inflammatory transcription control pathways, NFκB and AP-1, was found in AA subjects who also independently reported experiencing greater perceived racial discrimination as assessed through a 17-item Perceived Ethnic Discrimination Questionnaire—Community Version (34) when compared to EA subjects. These findings suggest that psychosocial stressors may play a role in the increased activation of pro-inflammatory pathways.

Existing studies have proposed several mechanisms in the intersection of psychological stress and AD. The first involves an association between the hypothalamus-pituitary-adrenal (HPA) axis and AD (35). AD has been linked to HPA axis alterations, which contribute to several downstream pathological changes in the skin. Adrenocorticotropic hormone (ACTH), which promotes glucocorticoid secretion, can create a negative feedback loop in the HPA axis in response to stress, and early life adversity can modulate the regulation of HPA axis through epigenetic modification of the glucocorticoid receptors in the skin (36). An important modulator in the HPA axis and AD pathogenesis is IL-18, a member of the IL-1 cytokine family implicated in various immune-mediated skin disease including AD, psoriasis, alopecia areata, dermatomyositis, and cutaneous lupus erythematous (37–41). ACTH can also activates caspase-1 and keratin 1, leading to keratinocyte production of IL-18 (42). In the absence of IL-12, IL-18 has been shown to modulate the Th_2 pathway, inducing expression of IL-4, IL-13, and IgE by basophils (43) The downstream effect of these expressed factors has been linked to AD pathogenesis. IL-13 is a prominent Th2 cytokine (6, 44); and serum IgE levels are elevated among AA patients with AD when compared to all other racial/ethnic groups (22). Lastly, IL-18 is thought to directly activate mast cells, leading to release of the enzyme chymase which cleaves pro-IL-18 and potentially accelerates the inflammatory response in AD lesions (45). Existing evidence shows higher clinical AD severity index score (SCORing Atopic Dermatitis or SCORAD) correlate with increased serum IL-18 concentration in AD patients (46). However, as far as the authors are aware, no studies have compared HPA modulation and IL-18 expression across different ethnic groups in AD patients.

Another possible mechanism for the intersection of stress with AD pathogenesis involves chronic psychological stressors that induce serum epinephrine, norepinephrine, and cortisol levels, triggering a shift to a Th_2 cytokine profile (47). Though this has not been substantiated in AD through measurement of serum levels, salivary cortisol level has been correlated with SCORAD index scores among patients with elevated stress levels (48). In addition, genetic variants of interferon regulatory factor 2 (IRF2), a protein with crucial roles in immune response, including the regulation of IFNγ and basophil expansion, as well as the transcription of gasdermin D, are associated with AD risk in both AA and EAs (49). Gasdermin D is a critical mediator of inflammatory pathologies, with its non-canonical inflammasome signaling pathway leading to proteolytic activation of IL-1B and IL-18. The pro-inflammatory cytokine, IL-1B, along with TNF-alpha induce expression of 11 beta-hydroxysteriod dehydrogenase, are critical enzymes involved in cortisol synthesis and hypothesized to modulate pro-inflammatory cytokine expression in keratinocytes (50). Our recent work using skin and 3D human skin equivalents (HSE) also demonstrates that skin from AA patients exhibits stronger inflammatory response when compared with that from EA patients. The differentially expressed genes (DEG) in AA skin include those that encode immunoglobulins and their receptors such as *FCER1G*; proinflammatory genes such as *TNF*, *IL-32*; and different EDC and keratin genes. By investigating the effect of TNF signaling on HSE, we further demonstrated enhanced TNF pro-inflammatory effects in AA HSE (51).

PSORIASIS

Epidemiology

Psoriasis is a chronic IMSC with variable prevalence across populations. It has a lower prevalence among AA patients when compared with EAs (0.22% to 1.9% in AA vs 1.28% to 3.6% in EA) in the United States (52). Nevertheless, AA patients have been found to have more extensive disease involvement and higher rate of comorbidities including diabetes, hypertension, and hyperlipidemia, after controlling for age and body mass index, when directly compared to EAs (53, 54). While erythematous plaques with thick overlying scale is characteristic of plaque psoriasis, AA patients often present with less conspicuous erythema and a higher degree of dyspigmentation (55), which can often take months to years to resolve. These pigmentary changes can often be of equal or greater concern to patients than the psoriasis itself and contributes to the report of increased disease severity, greater psychological impact, treatment dissatisfaction, and decreased quality of life among AA patients (9, 54). There is currently little consensus regarding gender differences in psoriasis (56), particularly for underrepresented ethnic groups.

We used the CDM insurance-claim database to review the associated demographic factors with 272,913 psoriatic patients with diagnosis between 2014 and 2018 (**Table 1**). Specifically, the impact of gender on psoriasis was found to be significantly different across ethnic groups, with EA and AA women having higher psoriasis diagnosis rates compared to men (OR=1.10, 1.14); in contrast, the gender effect was not significant in Hispanic patients (p=0.098), and within the Asian population, females had significantly lower rate (OR=0.84). Previous studies showed ambiguous results when estimating the prevalence of psoriasis in each gender, but overall the differences are very minimal (57). We also observed that psoriasis risk was significantly higher for people with obesity across different ethnic groups, and the association between clinical visits for psoriasis was significantly lower among patients with lower income.

Genetics

Like AD, psoriasis has a complex genetic architecture with >80 different disease susceptibility loci identified, the majority of which only contribute to a modest effect of disease association. The HLA-Cw6 is the most prominent disease-associated signal, with >4 OR being revealed (58), and multiple different work have also highlighted other independent signals in the MHC region (59, 60). Despite >10 years since the first GWAS for psoriasis was conducted, large-scale genetic studies for psoriasis have been exclusively been based on EA, Chinese, and Japanese populations (4, 7, 61–64). Up to now, only very limited small GWAS study on psoriasis is based on African ancestry (65). While this can be attributed by the lower incidence rate of psoriasis and the more complex design for GWAS in individuals of African ancestry, the lack of diversity in genetic research for psoriasis needs to be addressed in order to understand the disease heterogeneity and to facilitate the fine-mapping of ethnic-shared/unique causal variations.

Stress and Immunologic Parameters

Increased corticotropin-releasing hormone (CRH) from stress leads to increased serum cortisol levels and decreased brain derived neurotrophic factor (BDNF) (66). CRH, encoded by the *CRHR-1* gene, is a peptide hormone that is essential in the physiologic response to stress. Elevation in serum CRH levels is associated with exposure to stress in psoriatic and AD patients, providing a link between stress and the HPA in both conditions (67). In psoriatic patients, increased serum CRH and decreased skin CRHR-1 expression have also been linked to the induction of vascular endothelial growth factor (VEGF) release from mast cells (67). VEGF is a known growth factor involved in the pathogenesis of psoriatic lesions and it is therefore hypothesized that these pathways are linked through increased levels of CRH playing a role in the activation of mast cells that release VEGF (68). CRH has also been implicated in stimulating the production of IL-6 and IL-11, and the downregulation of IL-1B, IL-2, and IL-18 in keratinocytes (69, 70); and a previous work has found elevated serum cortisol level in psoriatic patients when compared to healthy controls under increased psychosocial stress (71). These authors proposed that psoriatic patients have a robust neuroendocrine response in the presence of acute stressors, increasing vulnerability to psoriatic activity. On the other hand, localized glucocorticoid deficiency in psoriatic skin is associated with epidermal differentiation and inflammatory response, and restoring glucocorticoid biosynthesis can normalize these processes (72). Topical glucocorticoid has been shown to be responded by AA skin in a stronger degree, with the response of AA skin associating with inflammation and metabolic disruptions while the genes responding to glucocorticoid in EA are associated with cell barrier modifications (73). Animal studies have demonstrated the reduction of brain-derived neurotrophic factor (BDNF), a protein with skin related functions in humans including the induction of apoptosis in basal keratinocytes, in response to acute stress (66, 74, 75). These findings suggest that psychosocial stress as a potential factor linking decreased BDNF levels in psoriatic patients. Nevertheless, there is still very limited study that describes the differential impact of stress on chemokine or cytokine expression among AA and EA patients with psoriasis, which requires attention and future investigations.

CONCLUSION

AD and psoriasis are two of the most common IMSCs that can have significant impact on the quality of life in patients, especially in racial/ethnic minorities. Individuals of African ancestry are more likely to develop AD in childhood, experience more severe disease, and can have atypical presentation in adulthood with greater involvement of the extensor surfaces. Though the prevalence of psoriasis is lower among AA patients, they can suffer from more extensive disease involvement, experience significant post-inflammatory changes, and report decreased quality of life. Previous studies have attempted to account for the differences in AD or psoriasis disease severity and prevalence through variable demographic and genetic components; however, as highlighted in a recent review, the differential severity of AD between AA and EA patients cannot solely be explained by association with genetic ancestry (31). Research on the ethnic heterogeneity of IMSCs should place more focus on the role of psychosocial stressors on inflammatory cytokine and chemokine expression and overall disease pathogenesis of these conditions. We review existing studies examining the role of psychological or psychosocial stress at the molecular level in AD and psoriasis, highlighting the role of the HPA axis and IL-18 in AD, CRH and BDNF in psoriasis, and cortisol levels in both.

AUTHOR CONTRIBUTIONS

LT and TJ planned and designed the work. QL, SS, and MP conducted the EHR analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Psoriasis Foundation (LT, MP, and JG), and awards from the National Institutes of Health (K01AR072129 to LT; 1P30AR075043 to LT, MP, and JG). MP, LT, TJ were also supported by the Dermatology Foundation.

REFERENCES

1. AAD. *Burden of Skin Disease* (2016). Available at: https://www.aad.org/member/clinical-quality/clinical-care/bsd.
2. Baurecht H, Hotze M, Brand S, Buning C, Cormican P, Corvin A, et al. Genome-Wide Comparative Analysis of Atopic Dermatitis and Psoriasis Gives Insight Into Opposing Genetic Mechanisms. *Am J Hum Genet* (2015) 96:104–20. doi: 10.1016/j.ajhg.2014.12.004
3. Li B, Tsoi LC, Swindell WR, Gudjonsson JE, Tejasvi T, Johnston A, et al. Transcriptome Analysis of Psoriasis in a Large Case-Control Sample: RNA-Seq Provides Insights Into Disease Mechanisms. *J Invest Dermatol* (2014) 134:1828–38. doi: 10.1038/jid.2014.28
4. Patrick MT, Stuart PE, Raja K, Gudjonsson JE, Tejasvi T, Yang J, et al. Genetic Signature to Provide Robust Risk Assessment of Psoriatic Arthritis Development in Psoriasis Patients. *Nat Commun* (2018) 9:4178. doi: 10.1038/s41467-018-06672-6
5. Patrick MT, Stuart PE, Zhang H, Zhao Q, Yin X, He K, et al. Causal Relationship and Shared Genetic Loci Between Psoriasis and Type 2 Diabetes Through Trans-Disease Meta-Analysis. *J Invest Dermatol* (2021) 141:1493–502. doi: 10.1016/j.jid.2020.11.025
6. Tsoi LC, Rodriguez E, Degenhardt F, Baurecht H, Wehkamp U, Volks N, et al. Atopic Dermatitis Is an IL-13 Dominant Disease With Greater Molecular Heterogeneity Compared to Psoriasis. *J Invest Dermatol* (2019) 139:1480–9. doi: 10.1016/j.jid.2018.12.018
7. Tsoi LC, Stuart PE, Tian C, Gudjonsson JE, Das S, Zawistowski M, et al. Large Scale Meta-Analysis Characterizes Genetic Architecture for Common Psoriasis Associated Variants. *Nat Commun* (2017) 8:15382. doi: 10.1038/ncomms15382
8. Uppala R, Tsoi LC, Harms PW, Wang B, Billi AC, Maverakis E, et al. "Autoinflammatory Psoriasis"-Genetics and Biology of Pustular Psoriasis. *Cell Mol Immunol* (2021) 18:307–17. doi: 10.1038/s41423-020-0519-3
9. Alexis AF, Blackcloud P. Psoriasis in Skin of Color: Epidemiology, Genetics, Clinical Presentation, and Treatment Nuances. *J Clin Aesthet Dermatol* (2014) 7:16–24. Retrieved from: https://jcadonline.com
10. Kaufman BP, Alexis AF. Psoriasis in Skin of Color: Insights Into the Epidemiology, Clinical Presentation, Genetics, Quality-Of-Life Impact, and

Treatment of Psoriasis in Non-White Racial/Ethnic Groups. *Am J Clin Dermatol* (2018) 19:405–23. doi: 10.1007/s40257-017-0332-7

11. Kaufman BP, Guttman-Yassky E, Alexis AF. Atopic Dermatitis in Diverse Racial and Ethnic Groups-Variations in Epidemiology, Genetics, Clinical Presentation and Treatment. *Exp Dermatol* (2018) 27:340–57. doi: 10.1111/exd.13514
12. Kantor R, Silverberg JI. Environmental Risk Factors and Their Role in the Management of Atopic Dermatitis. *Expert Rev Clin Immunol* (2017) 13:15–26. doi: 10.1080/1744666X.2016.1212660
13. Prescott SL, Larcombe DL, Logan AC, West C, Burks W, Caraballo L, et al. The Skin Microbiome: Impact of Modern Environments on Skin Ecology, Barrier Integrity, and Systemic Immune Programming. *World Allergy Organ J* (2017) 10:29. doi: 10.1186/s40413-017-0160-5
14. Vojdani A. A Potential Link Between Environmental Triggers and Autoimmunity. *Autoimmune Dis* (2014) 2014:437231. doi: 10.1155/2014/437231
15. Nevid JS, Rathus SA. *Psychology and the Challenges of Life: Adjustment in the New Millennium*. *10th*. Hoboken, New Jersey: Wiley (2007).
16. CDC. *Health*. Atlanta, Georgia:United States (2017). Available at: https://www.cdc.gov/nchs/data/hus/hus17.pdf.
17. Lee C, Ayers SL, Kronenfeld JJ. The Association Between Perceived Provider Discrimination, Healthcare Utilization and Health Status in Racial and Ethnic Minorities. *Ethn Dis* (2009) 19:330–7. Retrieved from: https://www.ethndis.org/edonline/index.php/ethndis
18. Peek ME, Wagner J, Tang H, Baker DC, Chin MH. Self-Reported Racial Discrimination in Health Care and Diabetes Outcomes. *Med Care* (2011) 49:618–25. doi: 10.1097/MLR.0b013e318215d925
19. Hanifin JM, Reed ML, Eczema P, Impact Working G. A Population-Based Survey of Eczema Prevalence in the United States. *Dermatitis* (2007) 18:82–91. doi: 10.2310/6620.2007.06034
20. Silverberg JI. Public Health Burden and Epidemiology of Atopic Dermatitis. *Dermatol Clinics* (2017) 35:283–9. doi: 10.1016/j.det.2017.02.002
21. Vachiramon V, Tey HL, Thompson AE, Yosipovitch G, Lotti R, Dallaglio K, et al. Atopic Dermatitis in African American Children: Addressing Unmet Needs of a Common Disease. *Pediatr Dermatol* (2012) 29:395–402. doi: 10.1111/j.1525-1470.2012.01740.x
22. Brunner PM, Guttman-Yassky E. Racial Differences in Atopic Dermatitis. *Ann Allergy Asthma Immunol* (2019) 122:449–55. doi: 10.1016/j.anai.2018.11.015
23. Tackett KJ, Jenkins F, Morrell DS, McShane DB, Burkhart CN. Structural Racism and Its Influence on the Severity of Atopic Dermatitis in African American Children. *Pediatr Dermatol* (2020) 37:142–6. doi: 10.1111/pde.14058
24. Gutowska-Owsiak D, de la Serna JB, Fritzsche M, Naeem A, Podobas EI, Leeming M, et al. Orchestrated Control of Filaggrin-Actin Scaffolds Underpins Cornification. *Cell Death Dis* (2018) 9:412. doi: 10.1038/s41419-018-0407-2
25. Furue M. Regulation of Filaggrin, Loricrin, and Involucrin by IL-4, IL-13, IL-17a, IL-22, AHR, and NRF2: Pathogenic Implications in Atopic Dermatitis. *Int J Mol Sci* (2020) 21:5382. doi: 10.3390/ijms21155382
26. O'Regan GM, Sandilands A, McLean WH, Irvine AD. Filaggrin in Atopic Dermatitis. *J Allergy Clin Immunol* (2008) 122:689–93. doi: 10.1016/j.jaci.2008.08.002
27. Margolis DJ, Apter AJ, Gupta J, Hoffstad O, Papadopoulos M, Campbell LE, et al. The Persistence of Atopic Dermatitis and Filaggrin (FLG) Mutations in a US Longitudinal Cohort. *J Allergy Clin Immunol* (2012) 130:912–7. doi: 10.1016/j.jaci.2012.07.008
28. Zhu Y, Mitra N, Feng Y, Tishkoff S, Hoffstad O, Margolis D. FLG Variation Differs Between European Americans and African Americans. *J Invest Dermatol* (2021) 141:1855–7. doi: 10.1016/j.jid.2020.12.022
29. Margolis DJ, Gupta J, Apter AJ, Ganguly T, Hoffstad O, Papadopoulos M, et al. Filaggrin-2 Variation Is Associated With More Persistent Atopic Dermatitis in African American Subjects. *J Allergy Clin Immunol* (2014) 133:784–9. doi: 10.1016/j.jaci.2013.09.015
30. Almoguera B, Vazquez L, Mentch F, March ME, Connolly JJ, Peissig PL, et al. Novel Locus for Atopic Dermatitis in African Americans and Replication in European Americans. *J Allergy Clin Immunol* (2019) 143:1229–31. doi: 10.1016/j.jaci.2018.10.038
31. Abuabara K, You Y, Margolis DJ, Hoffmann TJ, Risch N, Jorgenson E. Genetic Ancestry Does Not Explain Increased Atopic Dermatitis Susceptibility or Worse Disease Control Among African American Subjects in 2 Large US Cohorts. *J Allergy Clin Immunol* (2020) 145:192–8.e11. doi: 10.1016/j.jaci.2019.06.044
32. Wongvibulsin S, Sutaria N, Kannan S, Alphonse MP, Belzberg M, Williams KA, et al. Transcriptomic Analysis of Atopic Dermatitis in African Americans Is Characterized by Th2/Th17-Centered Cutaneous Immune Activation. *Sci Rep* (2021) 11:11175. doi: 10.1038/s41598-021-90105-w
33. Schmeer KK, Tarrence J. Racial-Ethnic Disparities in Inflammation: Evidence of Weathering in Childhood? *J Health Soc Behav* (2018) 59:411–28. doi: 10.1177/0022146518784592
34. Brondolo E, Kelly KP, Coakley V, Gordon T, Thompson S, Levy E, et al. The Perceived Ethnic Discrimination Questionnaire: Development and Preliminary Validation of a Community Version. *J Appl Soc Psychol* (2006) 35:335–65. doi: 10.1111/j.1559-1816.2005.tb02124.x
35. Lin TK, Zhong L, Santiago JL. Association Between Stress and the HPA Axis in the Atopic Dermatitis. *Int J Mol Sci* (2017) 18:2131. doi: 10.3390/ijms18102131
36. Liu PZ, Nusslock R. How Stress Gets Under the Skin: Early Life Adversity and Glucocorticoid Receptor Epigenetic Regulation. *Curr Genomics* (2018) 19:653–64. doi: 10.2174/1389202919666171228164350
37. Lebre MC, Antons JC, Kalinski P, Schuitemaker JH, van Capel TM, Kapsenberg ML, et al. Double-Stranded RNA-Exposed Human Keratinocytes Promote Th1 Responses by Inducing a Type-1 Polarized Phenotype in Dendritic Cells: Role of Keratinocyte-Derived Tumor Necrosis Factor Alpha, Type I Interferons, and Interleukin-18. *J Invest Dermatol* (2003) 120:990–7. doi: 10.1046/j.1523-1747.2003.12245.x
38. Nakanishi K, Yoshimoto T, Tsutsui H, Okamura H. Interleukin-18 Regulates Both Th1 and Th2 Responses. *Annu Rev Immunol* (2001) 19:423–74. doi: 10.1146/annurev.immunol.19.1.423
39. Tsoi LC, Gharaee-Kermani M, Berthier CC, Nault T, Hile GA, Estadt SN, et al. IL18-Containing 5-Gene Signature Distinguishes Histologically Identical Dermatomyositis and Lupus Erythematosus Skin Lesions. *JCI Insight* (2020) 5:e139558. doi: 10.1172/jci.insight.139558
40. Wang D, Drenker M, Eiz-Vesper B, Werfel T, Wittmann M. Evidence for a Pathogenetic Role of Interleukin-18 in Cutaneous Lupus Erythematosus. *Arthritis Rheum* (2008) 58:3205–15. doi: 10.1002/art.23868
41. Wittmann M, Macdonald A, Renne J. IL-18 and Skin Inflammation. *Autoimmun Rev* (2009) 9:45–8. doi: 10.1016/j.autrev.2009.03.003
42. Roth W, Kumar V, Beer HD, Richter M, Wohlenberg C, Reuter U, et al. Keratin 1 Maintains Skin Integrity and Participates in an Inflammatory Network in Skin Through Interleukin-18. *J Cell Sci* (2012) 125:5269–79. doi: 10.1242/jcs.116574
43. Yoshimoto T, Tsutsui H, Tominaga K, Hoshino K, Okamura H, Akira S, et al. IL-18, Although Antiallergic When Administered With IL-12, Stimulates IL-4 and Histamine Release by Basophils. *Proc Natl Acad Sci USA* (1999) 96:13962–6. doi: 10.1073/pnas.96.24.13962
44. Tsoi LC, Rodriguez E, Stolzl D, Wehkamp U, Sun J, Gerdes S, et al. Progression of Acute-to-Chronic Atopic Dermatitis Is Associated With Quantitative Rather Than Qualitative Changes in Cytokine Responses. *J Allergy Clin Immunol* (2019) 145:1406–15. doi: 10.1016/j.jaci.2019.11.047
45. Yoshimoto T, Mizutani H, Tsutsui H, Noben-Trauth N, Yamanaka K, Tanaka M, et al. IL-18 Induction of IgE: Dependence on CD4+ T Cells, IL-4 and STAT6. *Nat Immunol* (2000) 1:132–7. doi: 10.1038/77811
46. Zedan K, Rasheed Z, Farouk Y, Alzolibani AA, Bin Saif G, Ismail HA, et al. Immunoglobulin E, Interleukin-18 and Interleukin-12 in Patients With Atopic Dermatitis: Correlation With Disease Activity. *J Clin Diagn Res* (2015) 9:WC01–5. doi: 10.7860/JCDR/2015/12261.5742
47. Suarez AL, Feramisco JD, Koo J, Steinhoff M. Psychoneuroimmunology of Psychological Stress and Atopic Dermatitis: Pathophysiologic and Therapeutic Updates. *Acta Derm Venereol* (2012) 92:7–15. doi: 10.2340/00015555-1188
48. Mizawa M, Yamaguchi M, Ueda C, Makino T, Shimizu T. Stress Evaluation in Adult Patients With Atopic Dermatitis Using Salivary Cortisol. *BioMed Res Int* (2013) 2013:138027. doi: 10.1155/2013/138027
49. Gao PS, Leung DY, Rafaels NM, Boguniewicz M, Hand T, Gao L, et al. Genetic Variants in Interferon Regulatory Factor 2 (IRF2) Are Associated With Atopic

Dermatitis and Eczema Herpeticum. *J Invest Dermatol* (2012) 132:650–7. doi: 10.1038/jid.2011.374

50. Sollberger G, Choidas A, Burn GL, Habenberger P, Di Lucrezia R, Kordes S, et al. Gasdermin D Plays a Vital Role in the Generation of Neutrophil Extracellular Traps. *Sci Immunol* (2018) 3:eaar6689. doi: 10.1126/sciimmunol.aar6689
51. Klopot A, Baida G, Kel A, Tsoi LC, Perez White BE, Budunova I. Transcriptome Analysis Reveals Intrinsic Proinflammatory Signaling in Healthy African American Skin. *J Invest Dermatol* (2021) S0022-202X(21) 02400-3. doi: 10.1016/j.jid.2021.09.031
52. Rachakonda TD, Schupp CW, Armstrong AW. Psoriasis Prevalence Among Adults in the United States. *J Am Acad Dermatol* (2014) 70:512–6. doi: 10.1016/j.jaad.2013.11.013
53. Gelfand JM, Stern RS, Nijsten T, Feldman SR, Thomas J, Kist J, et al. The Prevalence of Psoriasis in African Americans: Results From a Population-Based Study. *J Am Acad Dermatol* (2005) 52:23–6. doi: 10.1016/j.jaad.2004.07.045
54. Kerr GS, Qaiyumi S, Richards J, Vahabzadeh-Monshie H, Kindred C, Whelton S, et al. Psoriasis and Psoriatic Arthritis in African-American Patients–the Need to Measure Disease Burden. *Clin Rheumatol* (2015) 34:1753–9. doi: 10.1007/s10067-014-2763-3
55. McMichael AJ, Vachiramon V, Guzman-Sanchez DA, Camacho F. Psoriasis in African-Americans: A Caregivers' Survey. *J Drugs Dermatol* (2012) 11:478–82. Retrieved from: https://jddonline.com
56. Iskandar IYK, Parisi R, Griffiths CEM, Ashcroft DM. Systematic Review Examining Changes Over Time and Variation in the Incidence and Prevalence of Psoriasis by Age and Gender. *Br J Dermatol* (2021) 184:243–58. doi: 10.1111/bjd.19169
57. Parisi R, Symmons DP, Griffiths CE, Ashcroft DM Identification and Management of Psoriasis and Associated ComorbidiTy (IMPACT) Project team, et al. Global Epidemiology of Psoriasis: A Systematic Review of Incidence and Prevalence. *J Invest Dermatol* (2013) 133:377–85. doi: 10.1038/jid.2012.339
58. Tsoi LC, Spain SL, Knight J, Ellinghaus E, Stuart PE, Capon F, et al. Identification of 15 New Psoriasis Susceptibility Loci Highlights the Role of Innate Immunity. *Nat Genet* (2012) 44:1341–8. doi: 10.1038/ng.2467
59. Fan X, Yang S, Huang W, Wang ZM, Sun LD, Liang YH, et al. Fine Mapping of the Psoriasis Susceptibility Locus PSORS1 Supports HLA-C as the Susceptibility Gene in the Han Chinese Population. *PloS Genet* (2008) 4: e1000038. doi: 10.1371/journal.pgen.1000038
60. Knight J, Spain SL, Capon F, Hayday A, Nestle FO, Clop A, et al. Conditional Analysis Identifies Three Novel Major Histocompatibility Complex Loci Associated With Psoriasis. *Hum Mol Genet* (2012) 21:5185–92. doi: 10.1093/hmg/dds344
61. Cargill M, Schrodi SJ, Chang M, Garcia VE, Brandon R, Callis KP, et al. A Large-Scale Genetic Association Study Confirms IL12B and Leads to the Identification of IL23R as Psoriasis-Risk Genes. *Am J Hum Genet* (2007) 80:273–90. doi: 10.1086/511051
62. Nair RP, Duffin KC, Helms C, Ding J, Stuart PE, Goldgar D, et al. Genome-Wide Scan Reveals Association of Psoriasis With IL-23 and NF-kappaB Pathways. *Nat Genet* (2009) 41:199–204. doi: 10.1038/ng.311
63. Ogawa K, Okada Y. The Current Landscape of Psoriasis Genetics in 2020. *J Dermatol Sci* (2020) 99:2–8. doi: 10.1016/j.jdermsci.2020.05.008
64. Yin X, Low HQ, Wang L, Li Y, Ellinghaus E, Han J, et al. Genome-Wide Meta-Analysis Identifies Multiple Novel Associations and Ethnic Heterogeneity of Psoriasis Susceptibility. *Nat Commun* (2015) 6:6916. doi: 10.1038/ncomms7916
65. Bejaoui Y, Witte M, Abdelhady M, Eldarouti M, Abdallah NMA, Elghzaly AA, et al. Genome-Wide Association Study of Psoriasis in an Egyptian Population. *Exp Dermatol* (2019) 28:623–7. doi: 10.1111/exd.13926
66. Bath KG, Schilit A, Lee FS. Stress Effects on BDNF Expression: Effects of Age, Sex, and Form of Stress. *Neuroscience* (2013) 239:149–56. doi: 10.1016/j.neuroscience.2013.01.074
67. Vasiadi M, Therianou A, Sideri K, Smyrnioti M, Sismanopoulos N, Delivanis DA, et al. Increased Serum CRH Levels With Decreased Skin CRHR-1 Gene Expression in Psoriasis and Atopic Dermatitis. *J Allergy Clin Immunol* (2012) 129:1410–3. doi: 10.1016/j.jaci.2012.01.041
68. Alexopoulos A, Chrousos GP. Stress-Related Skin Disorders. *Rev Endocr Metab Disord* (2016) 17:295–304. doi: 10.1007/s11154-016-9367-y
69. Park HJ, Kim HJ, Lee JH, Lee JY, Cho BK, Kang JS, et al. Corticotropin-Releasing Hormone (CRH) Downregulates Interleukin-18 Expression in Human HaCaT Keratinocytes by Activation of P38 Mitogen-Activated Protein Kinase (MAPK) Pathway. *J Invest Dermatol* (2005) 124:751–5. doi: 10.1111/j.0022-202X.2005.23656.x
70. Zbytek B, Pfeffer LM, Slominski AT. Corticotropin-Releasing Hormone Inhibits Nuclear factor-kappaB Pathway in Human HaCaT Keratinocytes. *J Invest Dermatol* (2003) 121:1496–9. doi: 10.1111/j.1523-1747.2003.12612.x
71. de Brouwer SJ, van Middendorp H, Stormink C, Kraaimaat FW, Sweep FC, de Jong EM, et al. The Psychophysiological Stress Response in Psoriasis and Rheumatoid Arthritis. *Br J Dermatol* (2014) 170:824–31. doi: 10.1111/bjd.12697
72. Sarkar MK, Kaplan N, Tsoi LC, Xing X, Liang Y, Swindell WR, et al. Endogenous Glucocorticoid Deficiency in Psoriasis Promotes Inflammation and Abnormal Differentiation. *J Invest Dermatol* (2017) 137:1474–83. doi: 10.1016/j.jid.2017.02.972
73. Lili LN, Klopot A, Readhead B, Baida G, Dudley JT, Budunova I. Transcriptomic Network Interactions in Human Skin Treated With Topical Glucocorticoid Clobetasol Propionate. *J Invest Dermatol* (2019) 139:2281–91. doi: 10.1016/j.jid.2019.04.021
74. Brunoni AR, Lotufo PA, Sabbag C, Goulart AC, Santos IS, Bensenor IM. Decreased Brain-Derived Neurotrophic Factor Plasma Levels in Psoriasis Patients. *Braz J Med Biol Res* (2015) 48:711–4. doi: 10.1590/1414-431x20154574
75. Truzzi F, Marconi A, Atzei P, Panza MC, Lotti R, Dallaglio K, et al. P75 Neurotrophin Receptor Mediates Apoptosis in Transit-Amplifying Cells and Its Overexpression Restores Cell Death in Psoriatic Keratinocytes. *Cell Death Differ* (2011) 18:948–58. doi: 10.1038/cdd.2010.162

Conflict of Interest: JG has served as a consultant to AbbVie, Eli Lilly, Almirall, Celgene, BMS, Janssen, Prometheus, TimberPharma, Galderma, Novatis, MiRagen, AnaptysBio and has received research support from AbbVie, SunPharma, Eli Lilly, Kyowa Kirin, Almirall, Celgene, BMS, Janssen, Prometheus, and TimberPharma. LT has received support from Janssen and Galderma.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

ORIGINAL RESEARCH
published: 13 May 2022
doi: 10.3389/fimmu.2022.838636

Single Cell Transcriptome and Surface Epitope Analysis of Ankylosing Spondylitis Facilitates Disease Classification by Machine Learning

Samuel Alber[1,2], Sugandh Kumar[2], Jared Liu[2], Zhi-Ming Huang[2], Diana Paez[3], Julie Hong[2], Hsin-Wen Chang[2], Tina Bhutani[2], Lianne S. Gensler[3] and Wilson Liao[2]*

[1] Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA, United States, [2] Department of Dermatology, University of California at San Francisco, San Francisco, CA, United States, [3] Division of Rheumatology, Department of Medicine, University of California at San Francisco, San Francisco, CA, United States

OPEN ACCESS

Edited by:
Xu-jie Zhou,
Peking University, China

Reviewed by:
Liza Konnikova,
Yale University, United States
Wei Zhou,
China Pharmaceutical University,
China

***Correspondence:**
Wilson Liao
wilson.liao@ucsf.edu

Specialty section:
This article was submitted to
Autoimmune and
Autoinflammatory Disorders,
a section of the journal
Frontiers in Immunology

Received: 18 December 2021
Accepted: 19 April 2022
Published: 13 May 2022

Citation:
Alber S, Kumar S, Liu J, Huang Z-M, Paez D, Hong J, Chang H-W, Bhutani T, Gensler LS and Liao W (2022) Single Cell Transcriptome and Surface Epitope Analysis of Ankylosing Spondylitis Facilitates Disease Classification by Machine Learning. Front. Immunol. 13:838636. doi: 10.3389/fimmu.2022.838636

Ankylosing spondylitis (AS) is an immune-mediated inflammatory disorder that primarily affects the axial skeleton, especially the sacroiliac joints and spine. This results in chronic back pain and, in extreme cases, ankylosis of the spine. Despite its debilitating effects, the pathogenesis of AS remains to be further elucidated. This study used single cell CITE-seq technology to analyze peripheral blood mononuclear cells (PBMCs) in AS and in healthy controls. We identified a number of molecular features associated with AS. CD52 was found to be overexpressed in both RNA and surface protein expression across several cell types in patients with AS. $CD16^+$ monocytes overexpressed *TNFSF10* and IL-18Rα in AS, while $CD8^+$ T_{EM} cells and natural killer cells overexpressed genes linked with cytotoxicity, including *GZMH, GZMB*, and *NKG7*. Tregs underexpressed CD39 in AS, suggesting reduced functionality. We identified an overrepresented NK cell subset in AS that overexpressed CD16, CD161, and CD38, as well as cytotoxic genes and pathways. Finally, we developed machine learning models derived from CITE-seq data for the classification of AS and achieved an Area Under the Receiver Operating Characteristic (AUROC) curve of > 0.95. In summary, CITE-seq identification of AS-associated genes and surface proteins in specific cell subsets informs our understanding of pathogenesis and potential new therapeutic targets, while providing new approaches for diagnosis *via* machine learning.

Keywords: ankylosing spondylitis, spondyloarthritis, single cell sequencing, CITE-seq, genomics, machine learning

INTRODUCTION

Affecting approximately 0.52-0.55% of the US population, ankylosing spondylitis (AS) is a chronic inflammatory disease that targets sacroiliac joints, spine, peripheral joints and entheseal attachment sites (1). In more severe cases, AS can cause fibrosis and calcification, resulting in ankylosis of the sacroiliac joints and spine (2). AS is part of a broader group of rheumatologic diseases commonly characterized by inflammatory back pain, enthesitis, and dactylitis known as spondyloarthritis (3). Extra-musculoskeletal manifestations of AS include acute anterior uveitis and psoriasis, and comorbidities include cardiovascular disease and osteoporosis (4–6). Additionally, it has been demonstrated that non-rheumatologists do not consider the diagnosis of AS in patients presenting with back pain, creating a delay in diagnosis and treatment (7). The most common method of ankylosing spondylitis diagnosis and classification is the modified New York Classification Criteria, which involves both radiological criterion, such as biliteral sacroiliitis grade ≥ II, and clinical criteria, such as limitation of chest expansion relative to values normal for age and sex (8).

Previous studies have pointed to the significance of genetic and immunological factors in AS. In particular, the major histocompatibility complex class I allele *HLA-B*27* was shown to be present in the majority of patients with AS, serving as a key biomarker for AS and determining a patient's susceptibility to the disease (9). Nevertheless, although the heritability of the susceptibility of AS is estimated to be around 90%, the contribution of *HLA-B*27* to this heritability is only roughly 20%, pointing to the presence of other genetic factors (10, 11). Other known genes contributing to AS include *ERAP1* and *ERAP2*. The IL-23/IL-17 axis has been shown to play a vital role in driving the inflammation behind AS.

Several cell types in the peripheral blood of patients with AS are thought to be involved in the pathogenesis of AS. Natural killer (NK) cells, while not expanded in AS (12), have been shown to respond to HLA-B27 *via* the KIR3DL1 receptor (13). $CD4^+$ T cells increase production of IP-10/CXCL10, which recruits Th1 cells that then amplify the inflammatory response *via* secretion of IFN-γ and TNF-α (14). Th17 cells have also been observed to play a key role in the pathogenesis of AS through the production of several inflammatory cytokines, such as IL-17 (15).

In this study, we used single-cell technology to help identify cellular composition differences as well as differentially expressed genes, proteins, and pathways in the peripheral blood mononuclear cells (PBMCs) of patients with AS. We utilized a multi-omic approach, surveying both transcriptome and cell surface proteins involved in this disease, and evaluated the diagnostic potential of these biomarkers using machine learning models to identify AS patients. To our knowledge, this is the first attempt to use machine learning and single cell transcriptome data to classify AS.

METHODS

Patient Recruitment and Sampling

Patients with ankylosing spondylitis (n=10; 6 male, 4 female) were enrolled from the rheumatology clinics at the University of California San Francisco (UCSF), with a board-certified rheumatologist confirming the clinical diagnosis of AS using the modified New York classification criteria. Nine of the ten AS subjects were not on any biologic therapy, while one AS subject was on ustekinumab for his concomitant Crohn's disease. Healthy controls (n=29), who did not have any inflammatory skin disease or autoimmune disease, were enrolled from the San Francisco Bay Area. All subjects gave written, informed consent under IRB approval 10-02830 from the University of California San Francisco. Detailed patient information is provided in **Supplementary Table 1**. Peripheral blood was collected from each subject in Vacutainer ACD tubes. PBMCs were isolated using a standard Ficoll method and stored in liquid nitrogen.

Sample and Library Preparation

Single Cell Libraries

500 μL thawed PBMCs from each subject were added to 10 mL EasySep (StemCell Technologies, Cat. 20144) and centrifuged (300G, 5 min, room temperature). Extracellular nucleic acids were digested by resuspending cell pellets in 1 mL of buffer made from 18 mL EasySep and 21 μL Benzonase Nuclease (MilliporeSigma, Cat. 70664) and incubating (15 min, room temperature). Nuclease-treated cell-suspensions were then filtered through a 40 μm Flowmi Cell Strainer (Bel-Art, Cat. H13680-0040), centrifuged (300G, 5 min, room temperature), and finally resuspended in 100 μL EasySep buffer. Cell counting was performed on 1:100 dilutions of final cell suspensions stained with 0.4% trypan blue using a Countess I FL Automated Cell Counter (Thermo Fisher Scientific).

Cell Surface Staining

Antibody staining of cell surface proteins was performed according to the Totalseq-A protocol (https://www.biolegend.com/en-us/protocols/totalseq-a-antibodies-and-cell-hashing-with-10x-single-cell-3-reagent-kit-v3-3-1-protocol) with modifications as follows. A pooled suspension containing 100,000 cells from at most 20 subjects at a time was centrifuged (300G, 5 min, 4°C) and resuspended in 100 μL Cell Staining Buffer (BioLegend, Cat. 420201) and incubated (10 min, 4°C) with 10 μL Human TruStain FcX^{TM} Fc Blocking Solution (BioLegend, Cat. 422301). Cells suspensions were then stained (30 min, 4°C) with 100 μL TotalSeq antibody cocktail (**Supplementary Table 2**) and divided into two 105 μL aliquots. Each aliquot was washed 3 times by resuspending in 15 mL Cell Staining Buffer and centrifuging (300G, 5 min, 4°C). Washed cells were then resuspended in 150 μL 10% FBS in PBS, recombined, and filtered again with a 40 μm Flowmi Cell Strainer. Cell viability was measured with 10 μL of filtered cells by adding 10 μL 0.4% Trypan Blue and manual counting with a hemocytometer. Cell density was adjusted to 2,500 cells/μL and run on the Chromium Controller (10X Genomics) using the Single Cell 3' v3.1 Assay (10X Genomics) with a target of 50,000 cells per reaction.

Library Preparation

Gene expression cDNA libraries were prepared using according to the manufacturer's instructions (https://assets.ctfassets.net/an68im79xiti/1eX2FPdpeCgnCJtw4fj9Hx/7cb84edaa9eca04b607f9193162994de/CG000204_ChromiumNextGEMSingleCell3_v3.1_Rev_D.pdf), with 12 cycles of PCR amplification. Libraries for antibody-derived tags (ADT) from feature barcoding antibodies were prepared by repeating size purification on the supernatant obtained from the prior size purification of gene expression cDNA libraries (Step 2.3.d in the manufacturer's instructions above), using 7:8 volumetric ratio of 2.0X SPRIselect reagent (Beckman Coulter, Cat# B23317) to sample. Indexing amplification was performed using Kapa Hifi HotStart ReadyMix (Kapa Biosystems, Cat# KK2601) and TruSeq Small RNA RPI primers (Illumina) with the following thermocycling conditions: (I) 98°C, 2 min; (II) 15 × (98°C, 20 sec; 60°C, 30 sec; 72°C, 20 sec); (III) 72°C, 5 min. Size purification was then repeated on amplified libraries using a 5:6 volumetric ratio of 1.2X SPRIselect reagent to sample. Libraries were quantified using a Bioanalyzer 2100 (Agilent) and sequenced on a Novaseq 6000 (Illumina).

Genotyping

DNA for genotyping was extracted from whole blood using the DNeasy blood and tissue kit (Qiagen, Cat. 69504). Extracted DNA was genotyped on the Affymetrix UK Biobank Axiom Array (ThermoFisher) using a GeneTitan Multi-Channel Instrument (Applied Biosystems).

Genotype Data Processing

SNPs were called using Analysis Power Tools 2.10.2.2 (Affymetrix, https://www.affymetrix.com/support/developer/) The resulting genotype vcfs were scanned with snpflip (https://github.com/biocore-ntnu/snpflip) using the GRCh37 build of the human genome reference sequence maintained by the University of California, Santa Cruz (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/bigZips/hg19.fa.gz) to identify reversed and ambiguous-stranded SNPs, which were flipped and removed (respectively) using Plink 1.90 (http://pngu.mgh.harvard.edu/purcell/plink/) (16), and the remaining sites were sorted using Plink 2.00a3LM (www.cog-genomics.org/plink/2.0/) (17). This SNP data was then augmented with additional sites imputed by the Michigan Imputation Server (https://imputationserver.sph.umich.edu) (1000G Phase 3 v5 GRCh37 reference panel, rsqFilter off, Eagle v2.4 phasing, EUR population). SNP positions were translated to GRCh38 coordinates using the 'LiftoverVcf' command of Picard 2.23.3 (http://broadinstitute.github.io/picard/). Finally, Vcftools 0.1.13 (18) was used to exclude non-exonic SNPs and SNPs with minor allele frequency < 0.05.

Single Cell Data Processing

Raw RNA and ADT fastqs for each Chromium library were respectively aligned to the GRCh38 human genome reference and the antibody-tag reference (**Supplementary Table 2**) using Cell Ranger 3.1.0 (10X Genomics) using default settings to obtain RNA and matched ADT (if available) count matrices for all barcodes representing non-empty droplets.

Cell Demultiplexing, Doublet Removal, and Annotation

Within each RNA count matrix, the subject of origin for all droplet barcodes was determined by using 'demuxlet' (19), as implemented in the 'popscle' suite (https://github.com/statgen/popscle) to imputation-augmented exonic SNP genotypes described above, and doublets detected between different individuals were excluded. The count matrices for each Chromium library were then loaded into R for analysis using the 'Seurat' 4.0.3 (20) R package, and the 'DoubletDecon' 1.1.6 R package (21) was used to further remove doublets formed by different cells within the same individual.

Annotations for each droplet barcode were determined by submitting raw RNA count matrices to Azimuth (https://azimuth.hubmapconsortium.org/) (20) for annotation with "celltype.l2" labels from the Human PBMC reference from Hao et al. (20).

Cell and Feature QC

We performed filtering of cells based on both RNA and ADT data by retaining cells with total RNA unique molecular identifiers (UMIs) between 500 and 10,000, total RNA features ≥ 200, percent mitochondrial and ribosomal protein reads in RNA ≤ 15% and 60% (respectively), total ADT features ≤ 260, and percent ADT reads mapping to 9 isotype control antibodies < 2%. In the RNA matrices of the resulting data, we further removed features (genes) with no detectable UMIs across the cells of all matrices. These matrices were finally merged together into a combined matrix of RNA data for all cells. In the ADT matrices, we further removed features corresponding to the 9 isotype controls and 15 features observed to have expression inconsistent with annotated cell types (**Supplementary Table 2**).

Intra-Cell Type Differential Feature Analysis and Clustering

To identify differentially expressed genes (DEGs) and proteins (DEPs), the Seurat object containing ADT and RNA expression from the QC'd dataset (see section 'Cell and feature QC' above) was subsetted by Azimuth-annotated cell type using 'SplitObject'. For each resulting Seurat object containing cells of a particular type, we performed normalization on RNA and ADT expression using SCTransform and CLR, again adjusting for total counts and total features in each cell (using the 'vars.to.regress' parameter). Differential gene expression between disease statuses as well as between clusters (see section 'Intra-cell type clustering') was then calculated on SCTransform-normalized counts using the negative binomial test (test.use = "negbinom" in Seurat). Genes with both Bonferroni-corrected p-value < 0.05 and absolute log fold change > 0.20 were considered significant. Differential protein analysis was performed similarly, except with the Wilcoxon test (test.use = "wilcox" in Seurat) on CLR-normalized, mean-centered and scaled ADT data (within the 'scale.data' slot of the Seurat object) only for cells with measured ADT data. The expression of DEGs and DEPs was compared between batches; DEGs and DEPs that had a clear overexpression in a small subset of the batches, such as *MTRNR2L12*, were filtered out. Pathway analysis was performed on differentially expressed genes *via* the 'gprofiler2' R package (22) against the Gene Ontology (GO), KEGG, and Reactome databases. For identification of transcription factors, gprofiler2 was also run against the TRANSFAC database. Afterwards, the p values returned by gprofiler2 were adjusted to FDR values for each database.

To identify phenotypic clusters within cell types, the RNA expression data for a cell type was first corrected for batch effects by first subsetting the raw count matrix by the cells within each sequencing batch. SCTransform was run individually for each count matrix, and the resulting SCT expression matrices were reintegrated into a single matrix (see section 'Data integration'). PCA was performed on the integrated SCT matrix, and the first 30 PCs were used to construct a shared nearest-neighbor network using the 'FindNeighbors' function. The network was then used to identify clusters with the 'FindClusters' function, using a resolution of 0.6. UMAPs were also generated from the first 30 PCs using the 'RunUMAP' function.

Data Integration

Integration of SCT expression data from two or more single-cell datasets was performed according to the Seurat data integration protocol (https://satijalab.org/seurat/articles/integration_introduction.html#performing-integration-on-datasets-normalized-with-sctransform-1). Briefly, 'SelectIntegrationFeatures' was used to select a common set of 3,000 genes most consistently variable among the individual SCT matrices, and 'PrepSCTIntegration' was then used to prepare reduced SCT expression matrices for just these genes. PCA was calculated for each reduced SCT matrix using 'RunPCA', and the first 50 principal components of this transformation were used to identify transcriptomically similar cells between each pair of reduced SCT matrices using 'FindIntegrationAnchors', with 'reduction' set to 'rpca'. Finally, an integrated SCT matrix was calculated using 'IntegrateData'.

Machine Learning Model Development

The input dataset for machine learning classification of AS and healthy subjects consisted of, for each subject, the means of sctransform-normalized, centered, and scaled expression of each feature in the set of cell-type-specific differentially expressed genes and proteins, calculated across that subject's cells within the corresponding cell types. These mean expression data for 39 subjects (29 healthy and 10 AS) were randomly assigned by a 50:50 ratio into training (healthy = 15 and AS = 5) and test (healthy = 14 and AS = 5) sets for ML model building and evaluation.

We first performed ensemble-based feature selection using the EFS-MI method (23) where subsets of the starting feature set predicted to be informative by four different ML algorithms (Feed Forward and Backward selection, Recursive RF, SVMRadial, and NNET) were combined and sorted by prediction potential classification rank. We selected the top twenty features to train nine ML algorithms based on linear, non-linear, and ensemble models provided by the 'caret' R package. Five-fold cross validation, repeated twice, was performed on the training set using each ML algorithm, and resulting models were evaluated on the test set. All essential tuning parameter were optimized with bootstrap = TRUE. For random forest (RF) models, the maximum number of tree splits in each step fixed a max_depth = (50, 80, 100, 150, 300), the maximum feature selected as auto (max_features = 'auto'), and error was minimized through impurity value (min_impurity_decrease = c(0.0, 0.02, 0.1, 0.5). Further, a minimum tree split per leaf in each step (min_samples_leaf = (1 to 10) while maximum generation of trees (n_estimator = 20) was considered, other parameters kept as a default for RF. For SVMRadial, we tuned cost and sigma factor for correct classification. In avNNet and Naïve Bayes, we used TRUE kernel, decays, and their size as a tuning parameter. Model performance and robustness were evaluated based on classification statistics that include accuracy, area under receiver operating characteristic curve (AUROC), specificity, sensitivity, F1 score (harmonic mean of precision and recall), and balanced accuracy (kappa).

To check for model bias due to potentially shared information between test and training subsets (which are derived from data normalized by 'sctransform' over cells from all subjects), we regenerated the input dataset using an alternative normalization approach that aggregates single-cell data only within the cells of each subject into a representative expression profile for each subject. Specifically, the expression value for a gene [or protein] feature for a given subject was calculated as

$$\ln\left(\frac{\text{feature counts across all cells in subject}}{\text{total counts across all cells in subject}} \times \text{ scaling factor}\right)$$

where the scaling factor was chosen to be near the maximum number of counts across all subjects (10^7 for RNA, 5×10^5 for ADT). Training and testing of models was performed as above to evaluate accuracy and kappa. AUROC was also calculated for select models trained using 10-fold cross validation with 10 repeats.

RESULTS

Identifying Significant Peripheral Blood Mononuclear Cell Types in AS

Single-cell sequencing of PBMCs from 10 patients with AS and 29 healthy controls yielded transcriptome profiles of 98,884 cells (19,348 cells from patients with AS, 79,536 from the healthy controls). Single-cell RNA and ADT analysis was conducted on 59,585 of these cells and just single-cell RNA analysis was conducted on the other 39,299 cells (**Supplementary Figure 1**). Reference-based categorization of AS and healthy PBMCs into 30 unique cell types (**Figure 1A**) revealed a significantly lower abundance of CD4$^+$ cytotoxic T cells and hematopoietic stem and progenitor cells (HSPCs) in AS subjects (**Figure 1B**).

Differentially Expressed Genes and Pathways Associated With AS in Circulating Immune Populations

Differentially expressed genes (DEGs) identified for each of the 30 identified cell types varied from 0 for several cell types to 88 for CD4$^+$ naïve T cells, resulting in 898 total DEGs across all cell types (**Supplementary Figure 2A**). Of these, 9 cell types with a high number of biologically significant DEGs are shown in **Figure 2**, with the rest found in **Supplementary Table 3**.

Biologically relevant genes were identified across several cell types (**Figure 2**). *CD52* was overexpressed in AS CD14$^+$

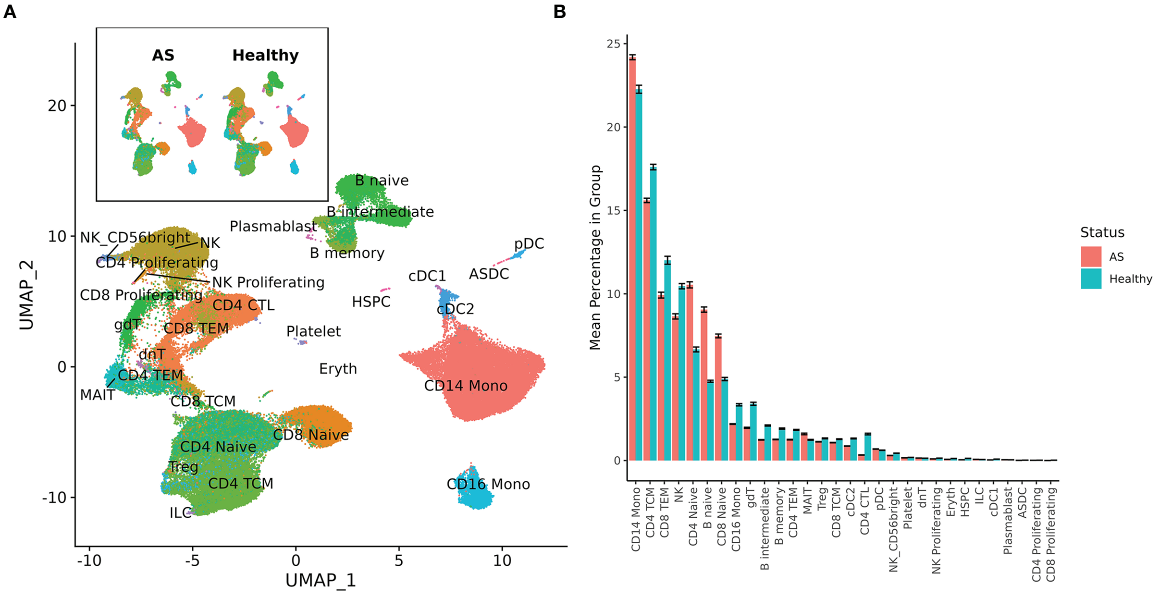


FIGURE 1 | Cell types among AS and healthy PBMCs. **(A)** UMAP of 30 cell types in AS and healthy PBMCs based on RNA expression. **(B)** Mean percentage of each cell type within the total PBMCs from each subject, averaged across AS and healthy cohorts. Error bars represent standard error of the mean. When tested for statistical significance using the Wilcoxon rank-sum test, CD4+ cytotoxic T cells and hematopoietic stem and progenitor cells were significantly underexpressed in AS.

monocytes, natural killer cells, and CD4$^+$ effector memory T (T$_{EM}$) cells. CD8$^+$ T$_{EM}$ cells and natural killer cells overexpressed genes linked with cytotoxicity, including *GZMH, GZMB*, and *NKG7*. *HLA-DRB5* was overexpressed in the CD14$^+$ monocytes, naïve B cells, memory B cells, and CD16$^+$ monocytes of AS patients. CD8$^+$ T$_{EM}$ cells also overexpressed *CMC1, CCL4*, and *CCL4L2* while natural killer cells overexpressed *S100A11*. We observed an upregulation of *CXCL8* in AS CD14$^+$ monocytes and an upregulation of *TNFSF10* in CD16$^+$ monocytes. CD4$^+$ T$_{CM}$ cells in patients with AS overexpressed *KLRB1*. Naïve B cells overexpressed *TCL1A* and *CXCR4*.

Comparison of inflammatory cytokines involved in AS, including *TNF, IL1B, IL17A, IFNG, IL23A, IL7R*, and *IL17F*, revealed comparable expression between AS and healthy subjects for each annotated cell type except for mucosal-associated invariant T (MAIT) cells, in which we observed a statistically significant decrease in *IL7R* expression in AS cells (**Supplementary Table 3**) that was also reflected in cell surface expression of IL-7Rα protein in the ADT data (**Supplementary Table 4**).

Proteomic Analysis Reveals Inflammatory Cell Surface Proteins in AS

Differentially expressed cell surface proteins (DEPs) were calculated for each cell type (0 – 24 features, **Supplementary Figure 2B**) between AS and healthy cells with ADT data measuring 258 cell surface proteins (**Figure 3**). Tregs in AS underexpressed CD39. CD14$^+$ monocytes and CD16$^+$ monocytes in AS overexpressed CD52. CD8$^+$ T$_{EM}$ cells overexpressed PD-1, KIR2DL2/L3 and KIR2DL1/S1/S3/S5. Natural killer cells in AS overexpressed CD16 but underexpressed CD94 and NKG2D. Memory B cells and CD16$^+$ monocytes overexpressed IL-18Rα. Memory B cells in AS also overexpressed CCR6 while naïve B cells overexpressed CD5 and CD74. Both CD4$^+$ T$_{EM}$ and MAIT cells in AS underexpressed IL7Rα. CD16$^+$ monocytes in AS overexpressed folate receptor β (FR-β).

De Novo Clustering Reveals an NK Subset Associated With AS

Due to the overexpression of CD16 in AS NK cells, the proportion of CD16$^+$ CD56^{dim} NK cells was compared between patients with AS and control patients *via* the Wilcoxon test, where it was found that CD16$^+$ CD56^{dim} NK cells were significantly overrepresented in patients with AS ($p = 0.006$; **Supplementary Figure 3A**). To investigate whether there was a subset in NK cells driving this overexpression of CD16, we performed *de novo* clustering on NK cells (**Supplementary Figure 3B**). A cluster was identified in NK cells that was statistically overrepresented in patients with AS (**Supplementary Figure 3C**). This subset overexpressed CD16, CD38, and CD161 on the ADT level and *SPON2, NKG7, FGFBP2, KLRB1*, and *MYOM2* on the RNA level (**Supplementary Table 5** and **Supplementary Figure 3D**). *De novo* clustering was also performed on CD8$^+$ T$_{EM}$ cells, naïve CD8$^+$ T cells, and CD14$^+$ monocytes, however no subsets in any of these cell types were statistically overrepresented in AS.

Gene Set Enrichment Analysis

To capture the relationships and shared pathways between differentially expressed genes in AS, we conducted functional enrichment analysis using gprofiler2 (**Supplementary Table 6**). Several pathways were significant at a nominal level ($p < 0.05$) but did not remain significant after FDR correction. CD14$^+$ monocytes were observed to upregulate pathways related to IL-4

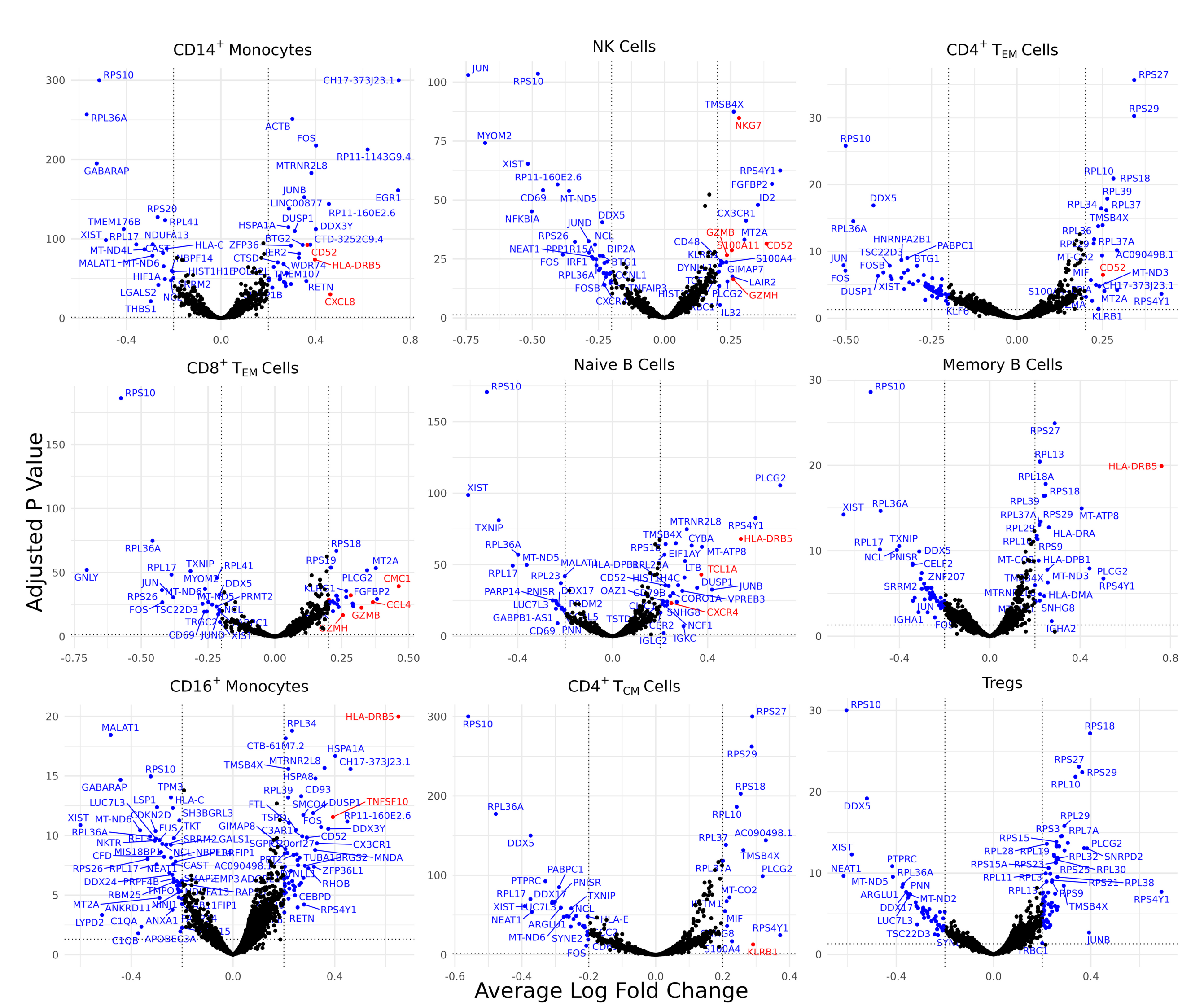


FIGURE 2 | Cell type-specific DE genes between AS and healthy subjects. Volcano plots showing the DEGs for 9 cell types based on a Bonferroni-adjusted p-value < 0.05 and absolute average log 2 fold change ≥ 0.20. Blue points represent statistically significant genes and genes mentioned in the paper are highlighted in red.

and IL-13 signaling, which were primarily constituted by *CXCL8, FOS*, and *JUNB*. Memory B cells upregulated pathways for Th17 cell differentiation, Th1/Th2 cell differentiation, and interferon-gamma-mediated signaling pathways.

To identify important transcription factors in our dataset, we conducted a separate gene set enrichment analysis using gprofiler2 and the TRANSFAC database. We found a set of transcription factors that were significant at the nominal level but did not remain significant after FDR correction. To help provide evidence for our results, we compared our identified transcription factors with a past study that used ATAC-seq on AS PBMCs (24). There, both our dataset and the ATAC-seq dataset found a statistically significant enrichment of GCM1, ETS1, ETV4, and ELF1 in CD4$^+$ T cells. The NK cell subset that was found to be overrepresented in AS during *de novo* clustering upregulated NK cell mediated cytotoxicity (p = 0.003); however, this pathway was no longer statistically significant after correction (FDR = 0.2).

Machine Learning Classification of AS

We next investigated the diagnostic potential of the cell type specific gene and protein expression differences we observed above by using these biomarkers to perform machine learning classification of AS and healthy subjects. Taking the mean normalized expression of each DE gene (**Supplementary Table 3**) or protein (**Supplementary Table 4**) across each subject's cells in the corresponding cell types, we performed ensemble feature selection (23) using four ML algorithms to identify an optimal subset of 18 features among the DE genes and 18 features among the DE proteins with the highest classification rate. Feature importance was generally higher among DE genes than proteins (**Figures 4A, B**), and applying this approach to the

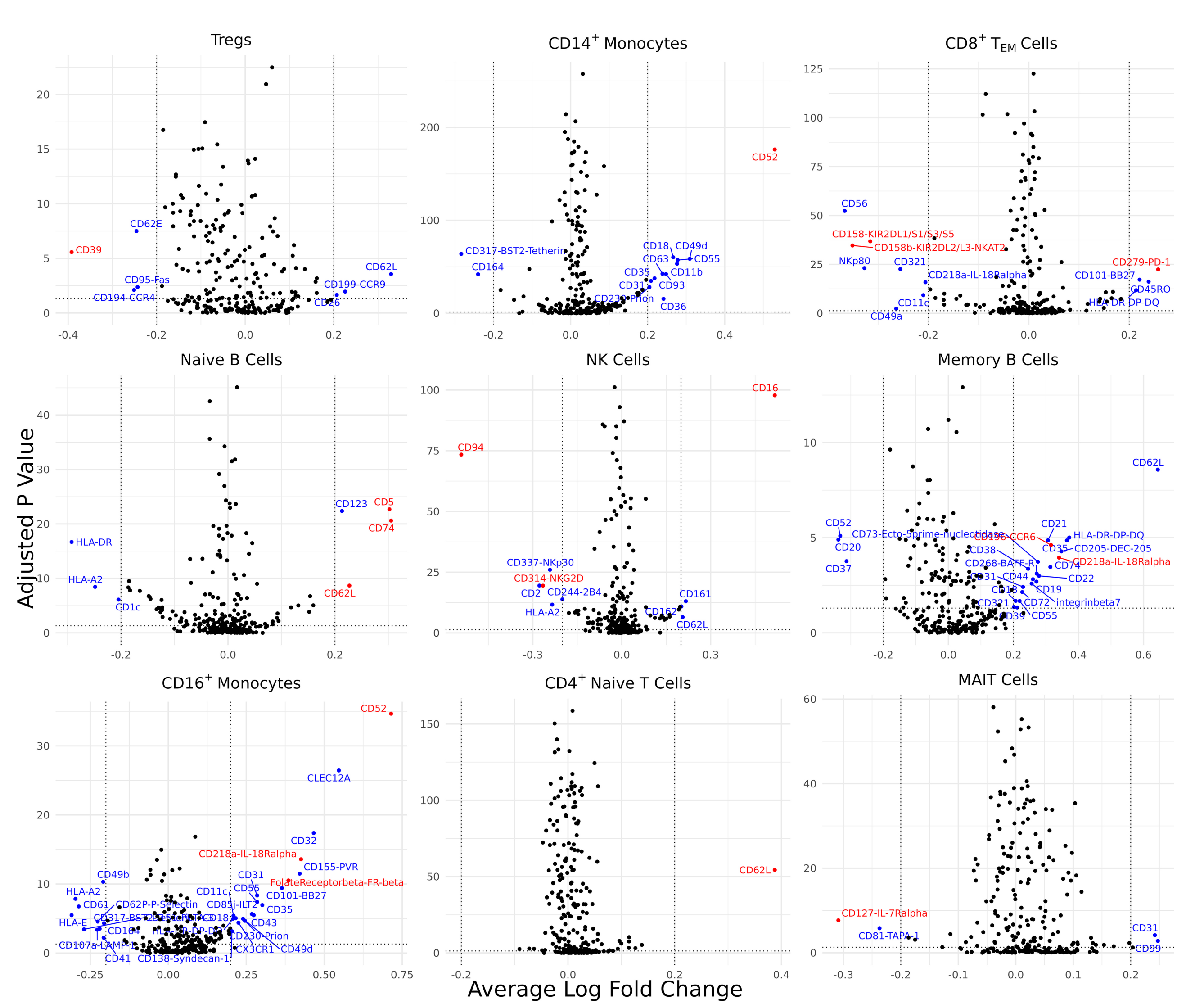


FIGURE 3 | Cell type-specific DE proteins between AS and healthy subjects. Volcano plots showing the DEPs for 9 cell types based on cells with measured ADT data; blue points represent statistically significant proteins and proteins mentioned in the paper are highlighted in red. DEPs were identified based on a Bonferroni-adjusted p-value < 0.05 and absolute average difference ≥ 0.20.

combined DE gene and DE protein sets yielded a set of 20 optimal features (5 DE proteins, 15 DE genes) with classification rates similar to that of DE proteins (**Figure 4C**). Nine ML classifiers trained on each of these optimized feature sets yielded an average accuracy of 0.63 – 0.90 for DE genes, 0.76 – 0.93 for DE proteins and 0.80 – 0.96 for the set of combined features (**Figures 4D–F**), and kappa ranged between 0.61 – 0.88 for all models. For DE genes, the three best-performing models, SVMRadial, RF, and Naïve Bayes, achieved AUROCs of 0.87, 0.97, and 1.00 (**Figure 4G**), respectively, while the corresponding best three for DE proteins, Naïve Bayes, RF, and avNNet, achieved AUROCs of 0.96, 1.00, and 1.00 (**Figure 4H**). Finally, the best three models for the combined DE gene and DE protein set, RF, Naïve Bayes, and avNNet, classified all test set subjects perfectly, achieving an AUROC of 1.00 (**Figure 4I**). Similar classification accuracy was achieved by models built using scaled, within-subject counts of each DEG or DEP (**Supplementary Figure 4**), indicating that model performance is not substantially explained by our normalization approach.

DISCUSSION

In this study, we performed a multi-omic analysis of ankylosing spondylitis, identifying transcriptomic and surface epitope changes associated with disease. Our single-cell approach also identified cell subsets that may contribute to pathogenesis, allowing for the further elucidation of key AS pathways.

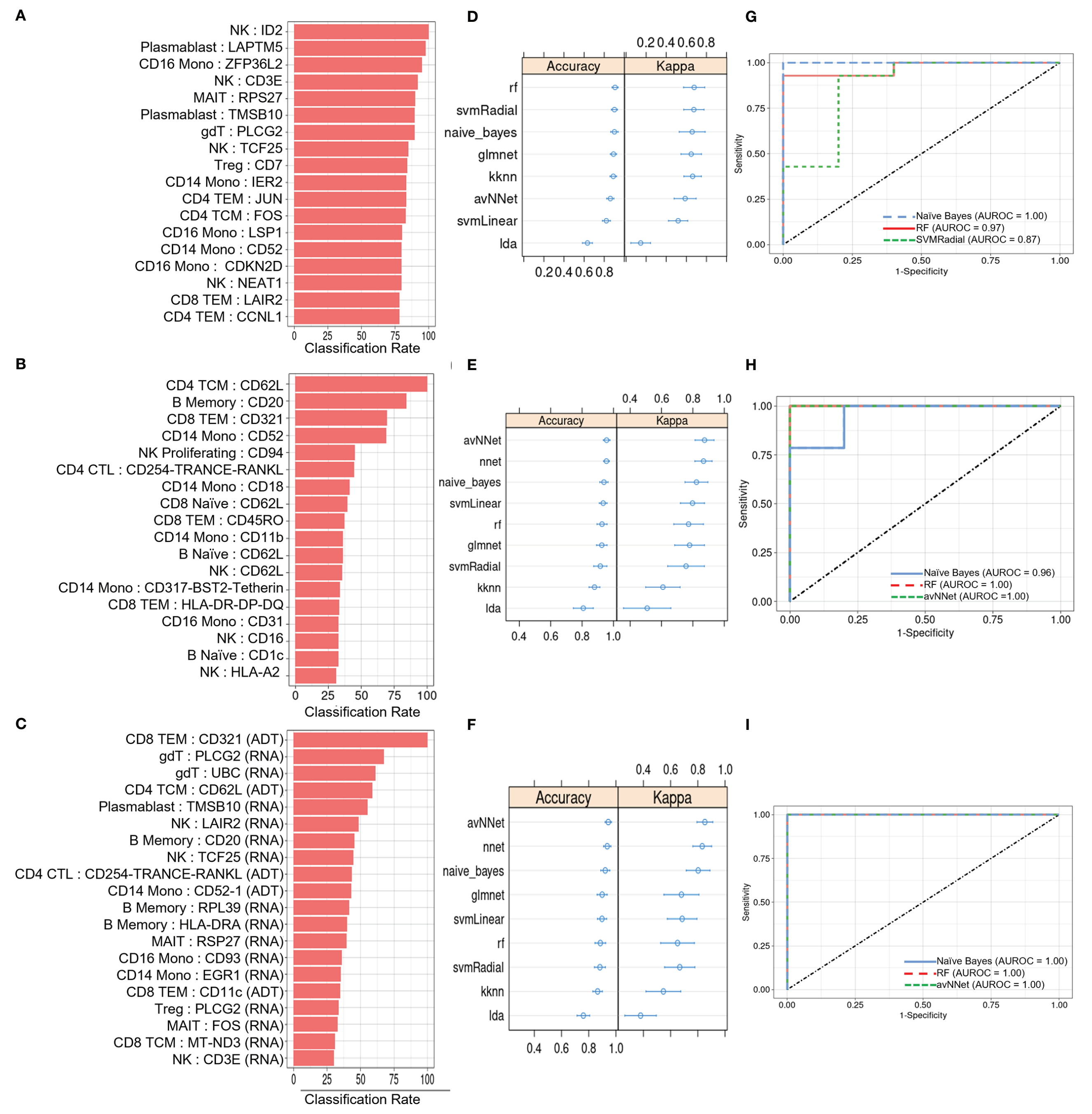


FIGURE 4 | ML classification of AS and healthy subjects. Classification rate of **(A)** top 18 DEGs, **(B)** top 18 DEPs, and **(C)** top 20 features from the set of combined DEGs and DEPs from ensemble feature selection, along with their respective **(D–F)** prediction accuracy and kappa using 9 ML models and **(G–I)** ROC curves based on test set classification. Error bars indicate 95% confidence interval.

AS PBMCs Overexpressed Genes and Proteins Linked With Inflammation

CD14$^+$ monocytes in patients with AS overexpressed *CXCL8*, which encodes IL-8. The increased production of IL-8 is correlated with AS (25). *CXCL8* is also known to induce *S100A11* expression (24), which is overexpressed in the AS NK cells in our dataset. Consequently, *CXCL8* in CD14$^+$ monocytes could be driving the observed increased production of IL-8 and other disease-related genes like *S100A11*. CD4$^+$ T$_{CM}$ cells in patients with AS overexpressed *KLRB1*, which is correlated with high tumor necrosis factor and interferon-γ co-expression potential in CD4$^+$ memory T cells. KLRB1$^+$

$CD4^+$ T_{CM} cells also have increased IL-17A production (26). As a result, increased KLRB1 expression in $CD4^+$ T_{CM} may upregulate inflammatory pathways and cytokines related to AS pathogenesis. CD39, an ecto-enzyme which converts extracellular ATP to extracellular adenosine, was significantly underexpressed in the Tregs of patients with AS. Tregs with low CD39 expression produce IL-17, contrary to their $CD39^+$ counterparts that suppress IL-17 production (27). Consequently, low CD39 expression on AS Tregs could be associated with limited functionality and the production of inflammatory IL-17.

Both memory B cells and $CD16^+$ monocytes in AS overexpressed IL-18Rα. IL-18 is a pro-inflammatory cytokine whose role in AS remains to be further elucidated with previous studies finding comparable IL-18 levels between healthy controls and AS patients (28). IL-18Rα also interacts with IL-37, which is significantly overexpressed in AS and may inhibit pro-inflammatory cytokine expression in AS PBMCs (29, 30). Consequently, increased IL-18Rα expression suggests the importance of cytokine signaling pathways in AS outside of the standard IL-23/IL-17 axis. On the other hand, both $CD4^+$ T_{EM} and MAIT cells in patients with AS underexpressed IL-7Rα relative to control patients, with MAIT cells also underexpressing *IL7R*. Our result contrasts with a MAIT-specific increase in IL-7R expression previously observed in AS patients that was associated with increased IL-17 expression (31), though further studies are needed to clarify the role of IL-7 signaling in AS.

Several inflammatory pathways were observed to be significant in AS at a nominal level, including the signaling of inflammatory pathways such as IL-4/IL-13 signaling (**Supplementary Table 5**). Many of these pathways involved the upregulation of *FOS* and *JUNB*, which are also linked to the abnormal expression of *NFKB* in AS $CD8^+$ T cells (32). Future studies are needed to follow up on these important inflammatory pathways.

Overall, our results suggest that a diverse set of cell types in the peripheral blood help drive the production of inflammatory cytokines. The identified surface proteins and genes could serve as potential therapeutic targets in AS.

Differential Genes and Proteins in AS Are Linked With Other Immune-Mediated Diseases

A subset of differentially expressed genes and proteins identified in our AS dataset are important in other known immune-mediated diseases. $CD8^+$ T_{EM} cells in AS overexpressed PD-1, which is a regulatory checkpoint inhibitor receptor for the immune system that has been proposed to play an important role in rheumatic disorders (33). Naïve B cells in AS overexpressed CD5. A similar result was found in a study on rheumatoid arthritis, which found that $CD5^+$ B cells may be involved in autoimmunity (34). $CD16^+$ monocytes in AS were seen to overexpress FR-β, which is part of a family of folate binding receptors. FR-β was upregulated in activated macrophages in the synovial tissue of patients with rheumatoid arthritis (35). Memory B cells overexpressed CCR6, which is a chemokine receptor with the ligand CCL20. CCR6 was seen to be overexpressed in the B cells of patients with systemic lupus erythematosus (SLE) (36). Naïve AS B cells also overexpressed CD74, which was similarly overexpressed in mice with an SLE phenotype (37). Naïve B cells in AS overexpressed *TCL1A*. A previous study has found that B cells in patients with Primary Sjögren's syndrome upregulated *TCL1A* (38). These genes and surface proteins could play a similar role in AS and suggest common treatment strategies across several types of immune-mediated diseases.

$CD14^+$ monocytes in patients with AS overexpressed CD52, a glycoprotein whose ligation results in T-cell activation and proliferation (39). Notably, CD52 is the therapeutic target of alemtuzumab, which is approved for the treatment of multiple sclerosis (40). Additionally, *CD52* was overexpressed across several AS cell types in our transcriptomic dataset, affirming the importance of CD52 at the transcriptome level and suggesting that *CD52* expression is upregulated in several cell types in the peripheral blood.

These results indicate that AS may share several differentially expressed genes and proteins with other immune-mediated diseases, indicating potential shared pathogenetic mechanisms and treatment strategies.

Cell Subsets, Genes, and Proteins Examined in Past AS Studies

In our study, there was a statistically significant overrepresentation of $CD16^+CD56^{dim}$ NK cells in patients with AS (**Supplementary Figure 2B**). Prior studies have shown that $CD16^+CD56^{dim}$ NK cells exhibit increased cytotoxic activity and are overexpressed in AS (41, 42). Although previous studies report conflicting observations on whether circulating NK cell abundance is altered in AS (12), by conducting *de novo* clustering on NK cells, we identified a subcluster that was overrepresented in AS (**Supplementary Figures 2D–F**). This subset had an overexpression of CD16, along with CD161 and CD38, which have been linked to cytotoxicity and pro-inflammatory NK cell subsets respectively (43, 44). Furthermore, a gene set enrichment analysis of this cluster revealed an upregulation of natural killer cell mediated cytotoxicity ($p = 0.03$). As a result, this cluster could be driving inflammation in AS and could consequently be a NK subset of interest for investigating NK cell activity in AS.

$CD8^+$ T_{EM} cells and NK cells in AS overexpressed genes related to cytotoxicity, including *GZMH, GZMB*, and *NKG7*. This result agrees with our finding that CD16 expression is increased in NK cells since high CD16 expression is linked with NK cell cytotoxicity (41). $CD8^+$ T_{EM} cells also overexpressed both *CCL4* and *CCL4L2*, which are inflammatory chemokines. These results provide further evidence for the increased cytotoxic activity of NK cells and $CD8^+$ T_{EM} cells in AS.

We compared transcription factors with p values below 0.05 against the data of a previous paper that used ATAC-seq on AS PBMCs to examine the role of transcription factors in AS (24). Our observed overexpression of GCM1, YY1, ETS1, ETV4, and ELF1 in $CD4^+$ T cells was confirmed in the ATAC-seq data, where all these transcription factors were also statistically significant. GCM1 had a particularly high log fold change in the ATAC-seq dataset, suggesting that this transcription factor may be particularly important.

CD8$^+$ T$_{EM}$ cells in AS overexpressed *CMC1*, variants of which are associated with AS (45). We also observed an overexpression of *TNFSF10* in the CD16$^+$ monocytes of AS patients relative to control patients. *TNFSF10* is part of the TNF superfamily and has been shown to be associated with AS pathogenesis (46). Naïve B cells in AS were observed to increase the expression of *CXCR4*. This gene has also been found to be upregulated in the hip synovial tissue of patients with AS (47).

CD14$^+$ monocytes, naïve B cells, and CD16$^+$ monocytes in AS displayed increased expression of *HLA-DRB5*. A previous study in the Chinese Han population found that increased DNA copy number of HLA-DQA1 but not HLA-DRB5 was associated with AS, though they did not measure transcription levels of these genes (48).

ML Classification of AS and Healthy Subjects

We found that AS-associated differences in cell-specific gene and surface protein expression could distinguish AS from healthy subjects, based on >0.95 AUROC achieved by several machine learning algorithms (**Supplementary Table 7**), though the general performance of these models may be limited by sample size (particularly for AS subjects). We nevertheless note that transcriptomic or cell surface protein expression of CD52 was consistently identified as an important feature in DEG, DEP, or DEG and DEP feature sets used for model training, which, given its biological significance as discussed above, may warrant further investigation as a diagnostic and therapeutic target.

Besides modest cohort size, other limitations of this study include the sampling of patients from a single center and the use of only molecular biomarkers for subject classification. Future multi-center studies can address these limitations by recruiting patients with AS and with back or joint pain from similar and unrelated diseases as well as by incorporating clinical and demographic data into the classification model.

Summary

This study has applied CITE-seq technology for the analysis of ankylosing spondylitis (AS), allowing for the important characterization of gene and cell surface proteins in AS. Numerous cell types overexpressed *CD52* on the transcriptomic and surface epitope level, which is involved in T-cell activation and is an important therapeutic target in other types of immune-mediated diseases. A pro-inflammatory NK cell subset was significantly overrepresented in AS that was characterized by high expression of CD16, CD161, and cytotoxic genes. This subset could be driving the overrepresentation of CD16$^+$ CD56^{dim} NK cells, a subset of NK cells with high cytotoxic activity, that was observed in our dataset and previous studies. *CD39* was underexpressed in Tregs, whose underexpression has been linked to IL-17 production and loss of functionality. CD14$^+$ monocytes in AS overexpressed *CXCL8*, which has been associated with increased inflammatory IL-8 expression.

Natural killer cells overexpressed cytotoxic genes along with *S100A11*, whose expression is induced by CXCL8. CD4$^+$ T$_{CM}$ cells in AS have a high expression of *KLRB1*, which is related to TNF and IFN-γ co-expression potential as well as IL-17A production. Memory B cells and CD16$^+$ monocytes overexpressed IL-18Rα, which interacts with the cytokines IL-18 and IL-37. CD5 was overexpressed in AS naïve B cells with CD5$^+$ B cells being known to be involved in autoimmunity.

Together, these results suggest cell type-specific changes both on the RNA level and on the surface protein level that may elucidate the pathogenesis of AS. The high classification rate of machine learning classifiers based on these gene and protein differences further indicates their potential as diagnostic biomarkers.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at https://www.ncbi.nlm.nih.gov/geo/, GSE194315.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of California, San Francisco IRB. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

WL and LG conceived and supervised the project. DP, JH, H-WC, TB, LG, and WL recruited study subjects and performed clinical annotation. Z-MH performed experimental procedures. SA, SK, JL, and WL performed data analysis. SA, SK, JL, LG, and WL wrote and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by grants from the National Psoriasis Foundation to WL. This study was also supported by PREMIER, a NIH/NIAMS P30 Center for the Advancement of Precision Medicine in Rheumatology at UCSF (P30AR070155). WL is supported by NIH grant R01AR078688. JL is supported by NIH grant T32AR007175.

ACKNOWLEDGMENTS

We would like to thank Laura Trupin, Jinoos Yazdany, and Mary Nakamura for their administrative support. We thank Chun Jimmie Ye for his technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.838636/full#supplementary-material

Supplementary Table 1 | Clinical and demographic characteristics of study subjects.

Supplementary Table 2 | TotalSeq Antibodies.

Supplementary Table 3 | Differentially expressed genes between AS and healthy controls.

Supplementary Table 4 | Differentially expressed proteins between AS and healthy controls.

Supplementary Table 5 | Gene and protein markers of *de novo* NK cell clusters.

Supplementary Table 6 | Gene set enrichment analysis between AS and healthy controls.

Supplementary Table 7 | ML classification performance.

REFERENCES

1. Reveille JD, Weisman MH. The Epidemiology of Back Pain, Axial Spondyloarthritis and HLA-B27 in the United States. *Am J Med Sci* (2013) 345:431–6. doi: 10.1097/maj.0b013e318294457f
2. Taurog JD, Chhabra A, Colbert RA. Ankylosing Spondylitis and Axial Spondyloarthritis. *N Engl J Med* (2016) 374:2563–74. doi: 10.1056/NEJMra1406182
3. Kataria RK, Brent LH. Spondyloarthropathies. *Am Fam Phys* (2004) 69:2853–60.
4. Rosenbaum JT. Uveitis in Spondyloarthritis Including Psoriatic Arthritis, Ankylosing Spondylitis, and Inflammatory Bowel Disease. *Clin Rheum* (2015) 34:999–1002. doi: 10.1007/s10067-015-2960-8
5. Van Praet L, den Bosch FE, Jacques P, Carron P, Jans L, Colman R, et al. Microscopic Gut Inflammation in Axial Spondyloarthritis: A Multiparametric Predictive Model. *Ann Rheum Dis* (2013) 72:414–7. doi: 10.1136/annrheumdis-2012-202135
6. Geusens P, Vosse D, van der Linden S. Osteoporosis and Vertebral Fractures in Ankylosing Spondylitis. *Curr Opin Rheum* (2007) 19:335–9. doi: 10.1097/BOR.0b013e328133f5b3
7. van Hoeven L, Luime J, Han H, Vergouwe Y, Weel A. Identifying Axial Spondyloarthritis in Dutch Primary Care Patients, Ages 20–45 Years, With Chronic Low Back Pain. *Arthritis Care Res (Hoboken)* (2014) 66:446–53. doi: 10.1002/acr.22180
8. Raychaudhuri SP, Deodhar A. The Classification and Diagnostic Criteria of Ankylosing Spondylitis. *J Autoimmun* (2014) 48–49:128–33. doi: 10.1016/j.jaut.2014.01.015
9. Brewerton DA, Hart FD, Nicholls A, Caffrey M, James DCO, Sturrock RD. Ankylosing Spondylitis and HL-A 27. *Lancet* (1973) 301:904–7. doi: 10.1016/S0140-6736(73)91360-3
10. Robinson PC, Brown MA. Genetics of Ankylosing Spondylitis. *Mol Immunol* (2014) 57:2–11. doi: 10.1016/j.molimm.2013.06.013
11. Breban M, Said-Nahal R, Hugot J-P, Miceli-Richard C. Familial and Genetic Aspects of Spondyloarthropathy. *Rheum Dis Clin* (2003) 29:575–94. doi: 10.1016/S0889-857X(03)00029-2
12. Liu D, Liu B, Lin C, Gu J. Imbalance of Peripheral Lymphocyte Subsets in Patients With Ankylosing Spondylitis: A Meta-Analysis. *Front Immunol* (2021) 12:696973. doi: 10.3389/fimmu.2021.696973
13. Chan AT, Kollnberger SD, Wedderburn LR, Bowness P. Expansion and Enhanced Survival of Natural Killer Cells Expressing the Killer Immunoglobulin-Like Receptor KIR3DL2 in Spondylarthritis. *Arthritis Rheum* (2005) 52:3586–95. doi: 10.1002/art.21395
14. Wang J, Zhao Q, Wang G, Yang C, Xu Y, Li Y, et al. Circulating Levels of Th1 and Th2 Chemokines in Patients With Ankylosing Spondylitis. *Cytokine* (2016) 81:10–4. doi: 10.1016/j.cyto.2016.01.012
15. Konya C, Paz Z, Apostolidis SA, Tsokos GC. Update on the Role of Interleukin 17 in Rheumatologic Autoimmune Diseases. *Cytokine* (2015) 75:207–15. doi: 10.1016/j.cyto.2015.01.003
16. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* (2007) 81:559–75. doi: 10.1086/519795
17. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets. *Gigascience* (2015) 4:7. doi: 10.1186/s13742-015-0047-8
18. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The Variant Call Format and VCFtools. *Bioinformatics* (2011) 27:2156–8. doi: 10.1093/bioinformatics/btr330
19. Kang HM, Subramaniam M, Targ S, Nguyen M, Maliskova L, McCarthy E, et al. Multiplexed Droplet Single-Cell RNA-Sequencing Using Natural Genetic Variation. *Nat Biotechnol* (2018) 36:89–94. doi: 10.1038/nbt.4042
20. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated Analysis of Multimodal Single-Cell Data. *Cell* (2021) 184:3573–87.e29. doi: 10.1016/j.cell.2021.04.048
21. DePasquale EAK, Schnell DJ, Van Camp P-J, Valiente-Alandí Í, Blaxall BC, Grimes HL, et al. DoubletDecon: Deconvoluting Doublets From Single-Cell RNA-Sequencing Data. *Cell Rep* (2019) 29:1718–27.e8. doi: 10.1016/j.celrep.2019.09.082
22. Kolberg L, Raudvere U, Kuzmin I, Vilo J, Peterson H. Gprofiler2 – an R Package for Gene List Functional Enrichment Analysis and Namespace Conversion Toolset G:Profiler. *F1000Research* (2020) 9:709. doi: 10.12688/f1000research.24956.2
23. Hoque N, Singh M, Bhattacharyya DK. EFS-MI: An Ensemble Feature Selection Method for Classification. *Complex Intell Syst* (2018) 4:105–18. doi: 10.1007/s40747-017-0060-x
24. Yu H, Wu H, Zheng F, Zhu C, Yin L, Dai W, et al. Gene-Regulatory Network Analysis of Ankylosing Spondylitis With a Single-Cell Chromatin Accessible Assay. *Sci Rep* (2020) 10:1–12. doi: 10.1038/s41598-020-76574-5
25. Azevedo VF, Faria-Neto JR, Stinghen A, Lorencetti PG, Miller WP, Gonçalves BP, et al. IL-8 But Not Other Biomarkers of Endothelial Damage is Associated With Disease Activity in Patients With Ankylosing Spondylitis Without Treatment With Anti-TNF Agents. *Rheum Int* (2013) 33:1779–83. doi: 10.1007/s00296-012-2631-x
26. Truong K-L, Schlickeiser S, Vogt K, Boës D, Stanko K, Appelt C, et al. Killer-Like Receptors and GPR56 Progressive Expression Defines Cytokine Production of Human CD4+ Memory T Cells. *Nat Commun* (2019) 10:2263. doi: 10.1038/s41467-019-10018-1
27. Fletcher JM, Lonergan R, Costelloe L, Kinsella K, Moran B, O'Farrelly C, et al. CD39+ Foxp3+ Regulatory T Cells Suppress Pathogenic Th17 Cells and are Impaired in Multiple Sclerosis. *J Immunol* (2009) 183:7602–10. doi: 10.4049/jimmunol.0901881
28. Przepiera-Bedzak H, Fischer K, Brzosko M. Serum Interleukin-18, Fetuin-A, Soluble Intercellular Adhesion Molecule-1, and Endothelin-1 in Ankylosing Spondylitis, Psoriatic Arthritis, and SAPHO Syndrome. *Int J Mol Sci* (2016) 17:1255. doi: 10.3390/ijms17081255
29. Wang X, Xu K, Chen S, Li Y, Li M. Role of Interleukin-37 in Inflammatory and Autoimmune Diseases. *Iran J Immunol* (2018) 15:165–74. doi: 10.22034/IJI.2018.39386
30. Chen B, Huang K, Ye L, Li Y, Zhang J, Zhang J, et al. Interleukin-37 is Increased in Ankylosing Spondylitis Patients and Associated With Disease Activity. *J Transl Med* (2015) 13:1–9. doi: 10.1186/s12967-015-0394-3
31. Gracey E, Qaiyum Z, Almaghlouth I, Lawson D, Karki S, Avvaru N, et al. IL-7 Primes IL-17 in Mucosal-Associated Invariant T (MAIT) Cells, Which Contribute to the Th17-Axis in Ankylosing Spondylitis. *Ann Rheum Dis* (2016) 75:2124–32. doi: 10.1136/annrheumdis-2015-208902
32. Xu H, Yu H, Liu L, Wu H, Zhang C, Cai W, et al. Integrative Single-Cell RNA-Seq and ATAC-Seq Analysis of Peripheral Mononuclear Cells in Patients

With Ankylosing Spondylitis. *Front Immunol* (2021) 12:760381. doi: 10.3389/fimmu.2021.760381

33. Zhang S, Wang L, Li M, Zhang F, Zeng X. The PD-1/PD-L Pathway in Rheumatic Diseases. *J Formos Med Assoc* (2021) 120:48–59. doi: 10.1016/j.jfma.2020.04.004
34. Shirai T, Hirose S, Okada T, Nishimura H. CD5+ B Cells in Autoimmune Disease and Lymphoid Malignancy. *Clin Immunol Immunopathol* (1991) 59:173–86. doi: 10.1016/0090-1229(91)90016-4
35. van der Heijden JW, Oerlemans R, Dijkmans BAC, Qi H, van der Laken CJ, Lems WF, et al. Folate Receptor β as a Potential Delivery Route for Novel Folate Antagonists to Macrophages in the Synovial Tissue of Rheumatoid Arthritis Patients. *Arthritis Rheum* (2009) 60:12–21. doi: 10.1002/art.24219
36. Lee AYS, Bannan JL, Adams MJ, Körner H. Expression of CCR6 on B Cells in Systemic Lupus Erythematosus Patients. *Clin Rheum* (2017) 36:1453–6. doi: 10.1007/s10067-017-3652-3
37. Lapter S, Ben-David H, Sharabi A, Zinger H, Telerman A, Gordin M, et al. A Role for the B-Cell CD74/macrophage Migration Inhibitory Factor Pathway in the Immunomodulation of Systemic Lupus Erythematosus by a Therapeutic Tolerogenic Peptide. *Immunology* (2011) 132:87–95. doi: 10.1111/j.1365-2567.2010.03342.x
38. Hong X, Meng S, Tang D, Wang T, Ding L, Yu H, et al. Single-Cell RNA Sequencing Reveals the Expansion of Cytotoxic CD4+ T Lymphocytes and a Landscape of Immune Cells in Primary Sjögren's Syndrome. *Front Immunol* (2021) 11:3688. doi: 10.3389/fimmu.2020.594658
39. Zhao Y, Su H, Shen X, Du J, Zhang X, Zhao Y. The Immunological Function of CD52 and its Targeting in Organ Transplantation. *Inflamm Res* (2017) 66:571–8. doi: 10.1007/s00011-017-1032-8
40. Hu Y, Turner MJ, Shields J, Gale MS, Hutto E, Roberts BL, et al. Investigation of the Mechanism of Action of Alemtuzumab in a Human CD52 Transgenic Mouse Model. *Immunology* (2009) 128:260–70. doi: 10.1111/j.1365-2567.2009.03115.x
41. Kucuksezer UC, Cetin EA, Esen F, Tahral I, Akdeniz N, Gelmez YM, et al. The Role of Natural Killer Cells in Autoimmune Diseases. *Front Immunol* (2021) 12:79. doi: 10.3389/fimmu.2021.622306
42. Mousavi T, Poormoghim H, Moradi M, Tajik N, Shahsavar F, Soofi M. Phenotypic Study of Natural Killer Cell Subsets in Ankylosing Spondylitis Patients. *Iran J Allergy Asthma Immunol* (2009) 8:193–8.
43. Piedra-Quintero ZL, Wilson Z, Nava P, Guerau-de-Arellano M. CD38: An Immunomodulatory Molecule in Inflammation and Autoimmunity. *Front Immunol* (2020) 11:597959. doi: 10.3389/fimmu.2020.597959
44. Kurioka A, Cosgrove C, Simoni Y, van Wilgenburg B, Geremia A, Björkander S, et al. CD161 Defines a Functionally Distinct Subset of Pro-Inflammatory Natural Killer Cells. *Front Immunol* (2018) 9:486. doi: 10.3389/fimmu.2018.00486
45. Ellinghaus D, Jostins L, Spain SL, Cortes A, Bethune J, Han B, et al. Analysis of Five Chronic Inflammatory Diseases Identifies 27 New Associations and Highlights Disease-Specific Patterns at Shared Loci. *Nat Genet* (2016) 48:510–8. doi: 10.1038/ng.3528
46. Zhu ZQ, Tang JS, Cao XJ. Transcriptome Network Analysis Reveals Potential Candidate Genes for Ankylosing Spondylitis. *Eur Rev Med Pharmacol Sci* (2013) 17:3178–85.
47. He C, Li D, Gao J, Li J, Liu Z, Xu W. Inhibition of CXCR4 Inhibits the Proliferation and Osteogenic Potential of Fibroblasts From Ankylosing Spondylitis *via* the Wnt/β-Catenin Pathway. *Mol Med Rep* (2019) 19:3237–46. doi: 10.3892/mmr.2019.9980
48. Wang J, Yang Y, Guo S, Chen Y, Yang C, Ji H, et al. Association Between Copy Number Variations of HLA-DQA1 and Ankylosing Spondylitis in the Chinese Han Population. *Genes Immun* (2013) 14:500–3. doi: 10.1038/gene.2013.46

Conflict of Interest: WL has received research grant funding from Abbvie, Amgen, Janssen, Leo, Novartis, Pfizer, Regeneron, and TRex Bio. LG has received research/grant support from Novartis, Pfizer and UCB and consulting fees from AbbVie, Eli Lilly, Gilead, Janssen, Novartis, Pfizer and UCB.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Immunology

ORIGINAL RESEARCH
published: 26 July 2022
doi: 10.3389/fimmu.2022.842651

Defining Patient-Level Molecular Heterogeneity in Psoriasis Vulgaris Based on Single-Cell Transcriptomics

Yale Liu[1,2,3], Hao Wang[4], Christopher Cook[2,3], Mark A. Taylor[3,5], Jeffrey P. North[3], Ashley Hailer[2,3], Yanhong Shou[6], Arsil Sadik[2], Esther Kim[7], Elizabeth Purdom[4], Jeffrey B. Cheng[2,3†] and Raymond J. Cho[3*†]*

[1] Department of Dermatology, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China, [2] Department of Dermatology, Veterans Affairs Medical Center, San Francisco, CA, United States, [3] Department of Dermatology, University of California, San Francisco, San Francisco, CA, United States, [4] Department of Statistics, University of California, Berkeley, Berkeley, CA, United States, [5] Clinical Research Centre, Medical University of Białystok, Białystok, Poland, [6] Department of Dermatology, Huashan Hospital, Fudan University, Shanghai, China, [7] Department of Plastic Surgery, University of California, San Francisco, San Francisco, CA, United States

Identifying genetic variation underlying human diseases establishes targets for therapeutic development and helps tailor treatments to individual patients. Large-scale transcriptomic profiling has extended the study of such molecular heterogeneity between patients to somatic tissues. However, the lower resolution of bulk RNA profiling, especially in a complex, composite tissue such as the skin, has limited its success. Here we demonstrate approaches to interrogate patient-level molecular variance in a chronic skin inflammatory disease, psoriasis vulgaris, leveraging single-cell RNA-sequencing of $CD45^+$ cells isolated from active lesions. Highly psoriasis-specific transcriptional abnormalities display greater than average inter-individual variance, nominating them as potential sources of clinical heterogeneity. We find that one of these chemokines, *CXCL13*, demonstrates significant correlation with severity of lesions within our patient series. Our analyses also establish that genes elevated in psoriatic skin-resident memory T cells are enriched for programs orchestrating chromatin and CDC42-dependent cytoskeleton remodeling, specific components of which are distinctly correlated with and against Th17 identity on a single-cell level. Collectively, these analyses describe systematic means to dissect cell type- and patient-level differences in cutaneous psoriasis using high-resolution transcriptional profiles of human inflammatory disease.

Keywords: single-cell RNA-sequencing, psoriasis vulgaris, heterogeneity, cytoskeleton, chromatin

OPEN ACCESS

Edited by:
Xu-jie Zhou,
Peking University, China

Reviewed by:
Anna Cogen,
University of Washington,
United States
Ali A. Zarrin,
TRex Bio, United States
Scott X. Atwood,
University of California, Irvine,
United States
Mona Ståhle,
Karolinska Institutet (KI), Sweden

***Correspondence:**
Raymond J. Cho
raymond.cho@ucsf.edu
Jeffrey B. Cheng
jeffrey.cheng@ucsf.edu

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Autoimmune and Autoinflammatory Disorders, a section of the journal Frontiers in Immunology

Received: 24 December 2021
Accepted: 19 April 2022
Published: 26 July 2022

Citation:
Liu Y, Wang H, Cook C, Taylor MA, North JP, Hailer A, Shou Y, Sadik A, Kim E, Purdom E, Cheng JB and Cho RJ (2022) Defining Patient-Level Molecular Heterogeneity in Psoriasis Vulgaris Based on Single-Cell Transcriptomics. Front. Immunol. 13:842651. doi: 10.3389/fimmu.2022.842651

INTRODUCTION

Individuals with psoriasis vulgaris broadly share cutaneous features such as erythema, micaceous scale, and induction at skin sites affected by friction. While the role of Th17 cell-produced cytokines such as *IL17F* and *IL26* in generating these phenotypes is well-established (1, 2), the distinctive morphology of these lesions suggests that a broad array of yet uncharacterized downstream effector genes are also specific to and shared by psoriatic lesions. Conversely, individual cases of psoriasis can markedly differ in presentation. Each patient develops lesions in distinct anatomic patterns, for example whether the scalp or intertriginous skin is involved, and lesional itch is also highly variable. These patterns of difference must reflect underlying molecular heterogeneity, potentially related to other clinical features such as involvement of other organ systems (*e.g.* psoriatic arthritis) or response to the many pathway-targeting agents now available for treatment. One well-established example is the involvement of germline *CARD14* variants in psoriasis patients with presentations overlapping with or including the disease state pityriasis rubra pilaris (3). However, many more yet undiscovered gene-based variances on the genetic and epigenetic level are likely to determine an individual patient's clinical state.

In the past, bulk RNA-sequencing of tissue obtained from lesional skin has been used to detect and define such commonalities and differences, enabling rough estimates of genetic variance in both psoriasis vulgaris and atopic dermatitis (4–6). Such approaches, however, conflate gene expression values from many different immune and stromal cell types, providing relatively crude estimates of genetic similarity and variance. The recent emergence of single-cell profiling technologies, such as single cell RNA sequencing (scRNA-seq) and Cellular Indexing of Transcriptomes and Epitopes (CITE-seq) (7), offers the ability to compare instances of chronic skin inflammatory disease with far greater resolution. We can now ask, for example, what molecular abnormalities are shared by effector immune cells in most psoriasis patients, regardless of clinical presentation? Such recurrent derangements might suggest treatment of psoriasis with existing drugs affecting those targets. Alternatively, certain molecular abnormalities are likely to be found in only a subset of individual cases, nominating them as candidates for specific targeted therapies.

To formally deconstruct discrete levels of molecular heterogeneity underlying cutaneous psoriatic inflammation, we analyzed data from a recent study profiling 8 psoriasis samples and 7 normal controls using single-cell RNAseq (scRNA-seq) and CITE-seq based on the 10X Genomics Chromium platform (8). We intended to develop and test approaches to scRNA-seq datasets profiling chronic inflammatory disease that could be practically and widely applied as similar datasets become published.

MATERIALS AND METHODS

Clinical Sample Acquisition

Patient recruitment and methods are detailed in our companion publication (8). Briefly, written informed consent was procured from donors providing both normal and psoriatic lesional skin under protocols approved by the University of California, San Francisco Institutional Review Board. Full thickness punch biopsies (6 mm) were obtained from psoriasis lesions; discards from abdominoplasties and mammoplasties were used as normal controls. All patients had not used topical immunosuppressives for at least 2 weeks before biopsy. All patients were naïve to targeted biologic medications or disease-modifying non-steroidal agents except for Patient 5, who was under systemic immunosuppression following a liver transplant. Clinical details of psoriasis samples in our series are described in **Supplementary Table 1**.

CD45$^+$ Immune Cell Isolation, Single-Cell RNA-seq and CITE-Seq Profiling, and Data Processing

Details of skin biopsy sample processing, CD45$^+$ immune cell isolation, 10X Genomics 3' scRNA-seq and CITE-seq library preparation, and data analysis are further described in a recent prior publication (8). Briefly, we initially performed high-resolution clustering and eliminated populations corresponding to non-immune and low quality cells (mitochondrial genes percentage <20%, 100 < nFeatures < 6000). With the remaining cells, we performed unsupervised clustering with the following in Seurat (15 harmonies to run UMAP() and 1.0 resolution for FindClusters()to obtain the final 20 clusters used in this analysis. Marker transcripts for each cluster were identified using the *FindAllMarkers* function in Seurat (results are in **Supplementary Table 2**). Cluster identities were then manually annotated based on canonical immune cell population markers.

Sample-Specific Differential Gene Expression Identification, Dispersion Score Calculation, and Metascape Analysis

We created pseudo-bulk counts for each patient for the cells that were mapped to CD45$^+$ cell subpopulations using the package *muscat* (9) in Bioconductor. The *muscat* method aggregates the single-cell data at the cluster-sample level to create pseudo-bulk data and then applies the methods of *edgeR* (10) to pseudo-bulk calculations to identify DEGs between normal and psoriasis samples (volcano plot, **Figure 2A**). To calculate dispersion values of the psoriasis samples, we applied the function estimateTagwiseDisp from the *edgeR* package in Bioconductor to the pseudo-bulk counts from the psoriasis samples. To identify abnormally elevated, functionally related gene sets (*e.g.* Gene Ontology (GO), Reactome) in Trm2, we applied the Metascape package (11, 12) to significant DEGs identified by FindMarkers() in comparison to grouped healthy controls ($p < 0.05$).

Normalization of T Cell Number Expressing Specific Immune Cell DEGs for Each Psoriasis Biopsy

Although all psoriasis biopsies were 0.6 cm in diameter, different proportions of isolated cells were scRNA-seq processed for each sample. To determine the number of T cells expressing each DEG in each biopsy, we took the assessed number of expressing T cells for a given DEG and adjusted by total number of CD45$^+$ cells obtained from each biopsy/total cell number processed in Seurat.

Statistical Correlation Analysis

Gene values were batch-corrected at the sample level using the CPCA method in the R package iCellR; missing gene values were independently imputed within inflamed and unflamed states of sample-aligned matrices using the PCA method in iCellR/run.impute. Resulting matrices were then used for the correlation matrix. Rstudio v.1.4.1717 and GraphPad Prism (version 8.0; GraphPad Software, La Jolla, California) were used for statistical analysis and heatmap generation. Pearson correlation coefficients were calculated for gene-gene comparisons using the R function cor(). Adjusted $p < 0.05$ was considered significant for Seurat-based analyses, while $p < 0.05$ was used for other analyses.

RESULTS

scRNA-Seq-Based Classification of Major T and Antigen-Presenting Cell Types Isolated From Psoriatic and Normal, Uninflamed Skin

We focused on 7 normal and 8 psoriasis samples from the Liu et al. study (8) (**Supplementary Table 1**). Diagnoses were based on clinical evaluation by a board-certified dermatologist and confirmed by formal histopathological reading. Six of eight patients were judged to have moderate to severe disease based on Psoriasis Area and Severity Index (PASI) scores and two (Patients 2 and 5) were in the mild range (**Supplementary Table 1**). The only patient known to be taking systemic immunosuppressive treatments within 4 weeks of biopsy was Patient 5, who was maintained daily on 4 mg of tacrolimus and 1250 mg of mycophenolate mofetil following a liver transplant. Normal controls were taken from discarded tissue obtained from mammoplasties and abdominoplasties.

Briefly, skin biopsies were enzymatically digested and flow sorted for live CD45$^+$ cells, which were then subjected to Chromium 3' single cell RNA-seq and CITE-seq protein epitope sequencing. Single-cell transcriptomic data was obtained from an average of ~5,200 single cells per sample after eliminating doublets, poor-quality, as well as non-immune cells. To classify cells, a graph-based clustering approach using Louvain community detection-based modularity optimization, available in the Seurat package, was utilized.

We obtained 20 cell types based on previously described unsupervised clustering approaches (8). Robust representation of each sample was observed (**Supplementary Data 1**). As shown in **Figure 1**, the most upregulated transcripts in each cluster (so-called marker genes) define a central memory cell population (*CD3D*$^+$/*CCR7*$^+$/*SELL*$^+$/*KLF2*$^+$) we call Tcm, as well as a migratory memory class Tmm (*CD3D*$^+$/*CCR7*$^+$/*SELL*$^-$). Based on expression of *ITGAE* (*CD103*), *CXCR6*, and *CD69*, we identified three resident memory populations (Trm1, Trm2, and Trm3). A *CD4*$^+$ regulatory T cell (Treg) population was noted based on the expression of *FOXP3, TIGIT, CTLA4, IL2RA (CD25)*, and *IKZF2* (Helios).

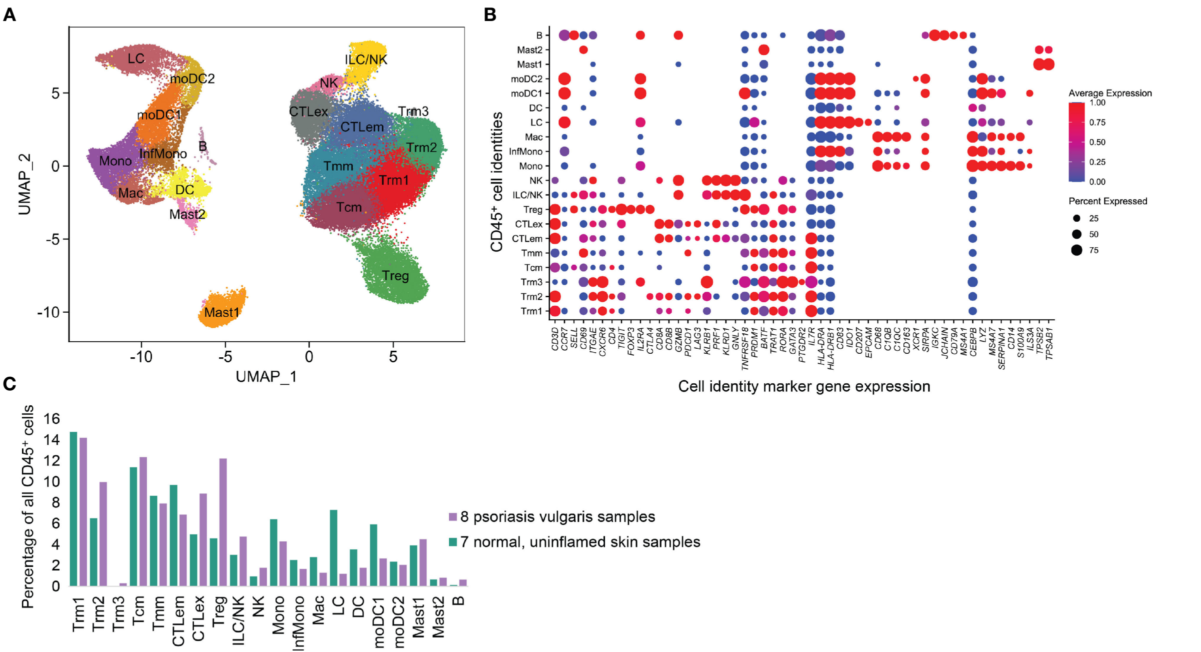


FIGURE 1 | CD45$^+$ immune cell types identified from 8 psoriasis vulgaris lesions and 7 normal skin samples. **(A)** UMAP representation of 11 T cell and 9 APC classes based on scRNA-seq transcriptional data, in which each point represents a single cell. **(B)** Expression of critical marker transcripts distinguishing immune cell classes. **(C)** Proportion of each immune class in total CD45$^+$ cell populations.

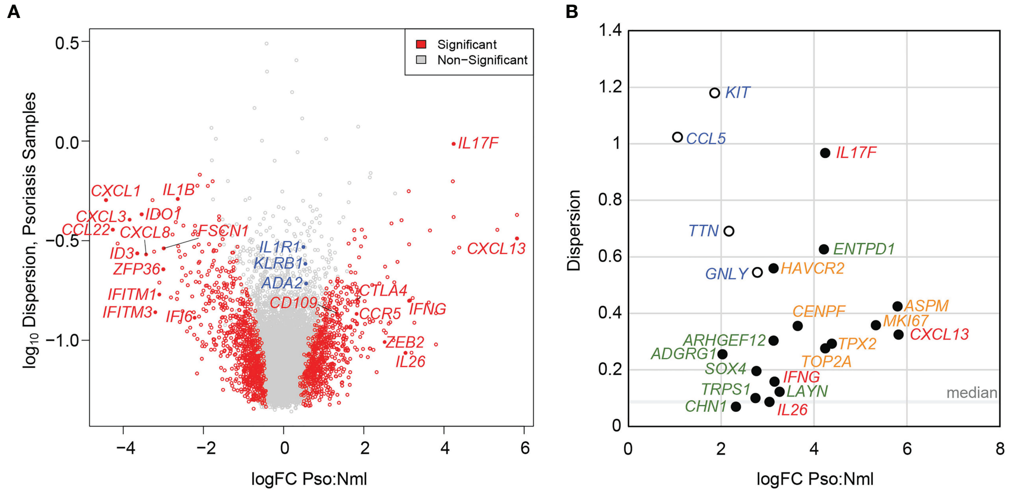


FIGURE 2 | Elevated patient level variance in psoriasis-specific skin-resident memory T cell (Trm2) DEGs. **(A)** Volcano plot showing psoriasis DEGs identified using a pseudo-bulk approach charted as a function of logFC difference from normal, uninflamed cells (*x*-axis) and the log of the dispersion score (a proxy for patient-level variation, *y*-axis). Significant DEGs are shown in red, non-significant DEGs in grey. Labelled in blue are immune activation genes with relatively high dispersion scores, which may have prevented them from reaching statistical significance. **(B)** LogFC (*x*-axis) and dispersion score (*y*-axis) shown for established pathogenic psoriatic cytokines (red), mitotic cell division transcripts (green), psoriasis-specific abnormalities not elevated in atopic dermatitis (orange), and as in **(A)**, immunologically activating DEGs with high end dispersion scores (blue).

Two cytotoxic (*CD8A*$^+$*CD8B*$^+$) T cell clusters expressing *CCL5, GZMB*, and *NKG7* were identified. One we annotated as cytotoxic effector memory cells *(*CTLem*)* due to expression of effector molecules including *TNFRSF18* and *CD96*, as well as resident markers *CD69* and *ITGAE.* Interestingly, the second cytotoxic T cell population was quantitatively enriched in the psoriasis vs. normal samples and contained elevated canonical exhaustion markers such as *PDCD1* and *LAG3*. Accordingly, this population was classified as exhausted T cells (CTLex). There were also two populations with high *KLRD1*$^+$, *GNLY*$^+$, *PRF1*$^+$, and *GZMB*$^+$ expression, one with high levels of the CD56 epitope by CITE-seq (NK cells) and the other defined as ILC/NK cells.

Antigen-presenting cell types (APCs) were also classified based on canonical markers. A macrophage population was enriched for *CD68*, *CEBPB*, and *FCER1G*, as well as complement transcripts *C1QB* and *C1QC* and the scavenger receptor *CD163* (Mac). We also examined four monocyte or monocyte-derived cell populations with elevated *MS4A7*, *LYZ*, and *SERPINA1*. There was an inflammatory monocyte (InfMono) population characterized by increased *IL1B* and *IL23A* and another cluster of classical monocytes (Mono) which expressed higher *S100A9* and *CD14*. Two of these clusters also expressed very high MHCII molecule levels (*HLA-DRA*, *HLA-DRB1*) and were identified as monocyte-derived DC (moDC1 and moDC2). A dendritic cell (DC) class (*HLA-DRA*$^+$) was enriched in *CLEC10A*. A population with *EPCAM*, and *CD207* was defined as Langerhans cells (LC). A small population comported with the B cell lineage, with high expression of *IGHG, IGHA, IGKC, JCHAIN, CD19*, and *MA4A2).* Two clusters of Mast cells (Mast) were distinguished by expression of *TPSB2* and *TPSAB1* (Mast1 and Mast2).

Psoriasis-Specific Transcriptional Abnormalities in Skin-Resident Memory T Cells Show High Patient-Level Variance

We next applied a pseudo-bulk method to identify differentially related genes (DEGs) that distinguished immune cell populations in our 8 psoriasis samples from 7 grouped healthy control biopsies. This approach aggregates scRNA-seq-derived gene counts for each cell subpopulation in each individual sample. Standard bulk mRNA-Seq computational approaches for differential expression were then applied, thereby allowing for patient-level variance to influence the significance of individual DEGs (9). One notable feature of our recent comparisons of psoriasis and other rash types such as atopic dermatitis is that the large majority of psoriasis-specific transcriptional changes are detected in Trm (8). For example, in the Tcm compartment, excluding mitochondrial and ribosomal transcripts, only *KLRB1*, *IL17R*, and *JUN* were expressed at greater than 0.5 logFC in psoriasis compared to normal samples. In Tregs, only *CPM*, *TNFRSF*, *CD7*, *FTH1*, *IL7R*, *MAGEH1*, *MAL*, *TBC1D4*, met these criteria. For APC classes, the far smaller number of cells captured in our CD45$^+$ cell-centric approach led to detection of even fewer highly specific DEGs.

Consistent with these recent findings, our pseudo-bulk analysis primarily detected upregulation of Th17 cytokines such as *IL17F* and *IL26*, as well as established psoriasis inflammatory markers such as *IFNG* and *CXCL13* (13), in a skin-resident memory T cell compartment (Trm2). Therefore, we mainly focused on this T cell class for further analysis. Overall, in Trm2 we identified 1,425 transcripts that distinguished psoriasis from healthy controls at a *p* value of < 0.05 (**Supplementary Table 3**).

Understanding patient-level variance, alongside fold-change magnitudes, is foundational to the conceptualization and use of disease biomarkers. Transcripts that distinguish psoriatic from normal tissue at higher log-fold change and relatively low patient-specific variance may perform well in broad screening efforts. Conversely, DEGs with high patient-level variance should be investigated as possible sources of phenotypic variance between affected individuals. To assess the contribution of patient-level variation to DEG identification in the skin-resident memory T cell population Trm2, we calculated dispersion values using *edgeR* (14) across our dataset. Lower dispersion scores correlate with lesser patient-level variation, which increases the significance of a pseudo-bulk-identified DEGs at a given logFC. **Figure 2A** plots logFC (*x*-axis) and dispersion score (*y*-axis), with transcripts with $p < 0.05$ adjusted value shown in red. In addition to the Th17 cytokines noted above, this representation shows significant elevation of immune activation markers such as *CTLA4*, *CCR5*, *CD109*, and *ZEB2*. We also saw clear suppression of other inflammatory pathways, including the interferon signaling genes *IFITM1*, *IFITM3*, and *IFI6* and the chemokines *CXCL3* and *CXCL8*. This representation also illustrates how greater patient-level variation for a given DEG (higher dispersion score along the *y*-axis) decreases its significance. For example, *IL1R1*, implicated in licensing Th17 cytokine production (15), *ADA2*, an adenosine deaminase central to T cell maturation (16), and the psoriasis-associated CD161 receptor gene *KLRB1* (17) show psoriasis-specific elevation in the logFC 0.5 range, but Log_{10} dispersion scores of greater than -1, likely contributing to their failure to reach statistical significance in comparison to healthy control skin-residency T cells (annotated in blue in **Figure 2A**, data in **Supplementary Table 3**).

We more closely examined inter-individual variance in psoriasis DEGs that were identified in the prior analysis as elevated not only relative to normal controls, but also to atopic dermatitis samples, indicating greater disease-specificity (8). Notably, many of these genes showed dispersion scores greater than the median of 0.086 (**Figure 2B**). In fact, for the six psoriasis DEGs with a logFC > 3 and significantly elevated compared to atopic dermatitis, the average dispersion score was 0.484 with a standard deviation of 0.273. In addition to the cytokines noted above such as *IL17F* (0.967), and *CXCL13* (0.325), this set contained identified psoriasis-specific genes with less established functional roles, such as *ARHGEF12* (0.303), *ENTPD1* (0.628), *LAYN* (0.122) and *HAVCR2* (0.560).

Cell cycle transcripts, which are elevated in both psoriasis and atopic dermatitis Trm2 compared to healthy controls, also show higher than median dispersion scores, including *MKI67* (0.358), *TOP2A* (0.276), and *CENPF* (0.356). Similar to **Figure 2A**, **Figure 2B** displays examples of psoriasis-implicated genes whose expression is elevated in Trm2, but whose high inter-individual variance reduces their overall significance level (*i.e. KIT*, *CCL5*, *TTN*, and *GNLY*, blue, open circles).

CXCL13 and *CD84* Expression in Cutaneous T Cells Corresponds With Lesional Psoriasis Severity

We were next curious to understand if expression of psoriasis-specific immune DEGs correlated with clinical features such as PASI score. Such relationships might further narrow the search for genetic factors influencing clinical heterogeneity in psoriasis. We chose 16 established immune activation genes from our Trm2 DEGs including *IL17F*, *CXCL13*, *IL26*, *CCR5*, and *CD82* (**Supplementary Table 4**) and quantified the T cells in each sample that detectably expressed each. We normalized these cell numbers between patients by bioinformatically deducing the total number of such cells existing in each sample, based on the total number of $CD45^+$ cells obtained from each biopsy, as well as the total number of scRNA-seq profiled cells processed in Seurat (Materials and Methods, **Supplementary Table 4**). To generate accompanying measures of clinical severity, we reasoned that the phenotype of a biopsied and molecularly profiled lesion would be best represented by summing its individual Erythema, Induration, and Desquamation PASI descriptors, rather than the overall patient score, and derived such a lesion-specific severity score for each sample (**Supplementary Table 1**).

We then assessed Spearman correlation of T cell expression of all 16 immune cell DEGs with lesion-specific severity score. Three of these genes correlated strongly with lesion-specific scores: a single gene coefficient of 0.851 for *CXCL13*, and *IKZF4* ($p = 7.3 \text{ x } 10^{-3}$) and 0.801 for *CD84* ($p = 1.7 \text{ x } 10^{-2}$) (Bonferroni unadjusted, eight selected genes displayed in **Figure 3**; **Supplementary Table 5**). When the patient-level PASI score was used as an alternative comparator, none of the 16 immune genes showed significant correlations at unadjusted *p* values.

Psoriatic $CD45^+$ Cells Show Programmatic Activation of Mitotic Cell Division, Chromatin Remodeling, CDC42 Signaling, and Leukocyte Activation

We next asked how functionally related groups of genes activated during psoriatic inflammation might vary in expression from patient to patient. We first applied the Metascape analysis package to detect overrepresentation of Gene Ontology and Reactome functional categories in the 662 genes significantly elevated (logFC > 0.4) in psoriatic skin-resident memory cells (Trm2), compared to healthy, controls, identifying 316 functional categories with a log (*q* value) < -2 (**Supplementary Table 6**; **Supplementary Data 2**). Statistically significant functional classes, included expected categories such as mitotic cell division and leukocyte activation (21 members, log (*q* value) < -9.97), but also highlighted the role of cytoskeletal reorganization (CDC42 signaling) and chromatin remodeling (**Figure 4A**). For example, *ARHGEF12* selectively regulates RhoA subfamily GTPases to coordinate cell migration and invasion (19), while *PAK2* influences actin cytoskeleton reorganization (20). *DOCK8* deficiencies impair immune cell migration in both the innate and adaptive immune system (21). Changes in psoriatic Trm also include elevated transcripts levels of the linker histone *H1FX* (22), histone chaperone *NAP1L4* (23), and the chromatin-modifying enzyme *SMARCA5* (24). **Figure 4B** globally displays psoriasis Trm2 abnormalities in these four programs on a per-patient level.

Considerable patient-specific fluctuations in these functionally related gene sets were easily appreciable. Most obviously, patient 5, the lone patient with psoriasis who was under systemic immunosuppression (mycophenolate and

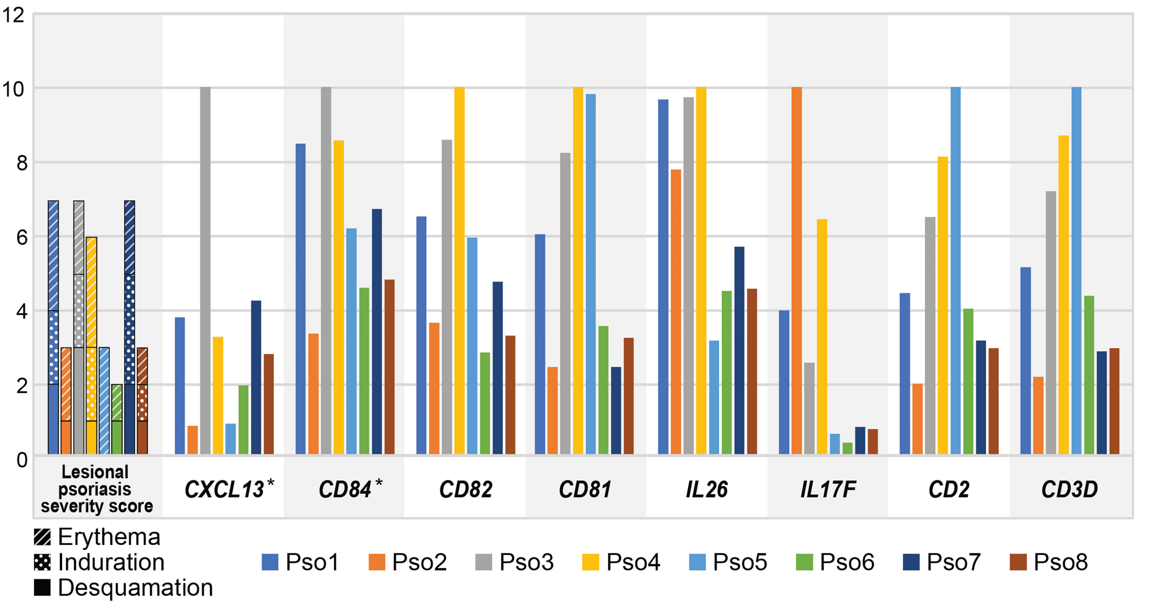


FIGURE 3 | Inter-individual variation in lesional psoriasis severity score parallels that of *CXCL13* and *CD84*. Leftmost graph shows severity scores for biopsied lesion for each patient. Subsequent graphs display deduced number of cells with positive expression for each gene, as a percentage of the maximum number of positive cells in any sample, multiplied by 1 x 10^3 (*i.e.* normalized to 10). Significant correlations for *CXCL13* and *CD84* are denoted by asterisks.

cyclosporine for a liver transplant) showed substantial attenuation of all these programs, corresponding to the lowest lesional psoriasis severity score (**Supplementary Table 1**). We sought to systematically assess these correlations between expression and phenotype, first averaging transcriptional log2FC for all genes in each of the four individual functional programs. Average scores for all four programs showed positive correlation with lesional severity score: CDC42 cytoskeletal reorganization at a Spearman rho value of 0.57, cell division at 0.55, chromatin reorganization at 0.48, and leukocyte activation at 0.36. None of these associations reached statistical significance, likely a factor of our limited sample size. However,

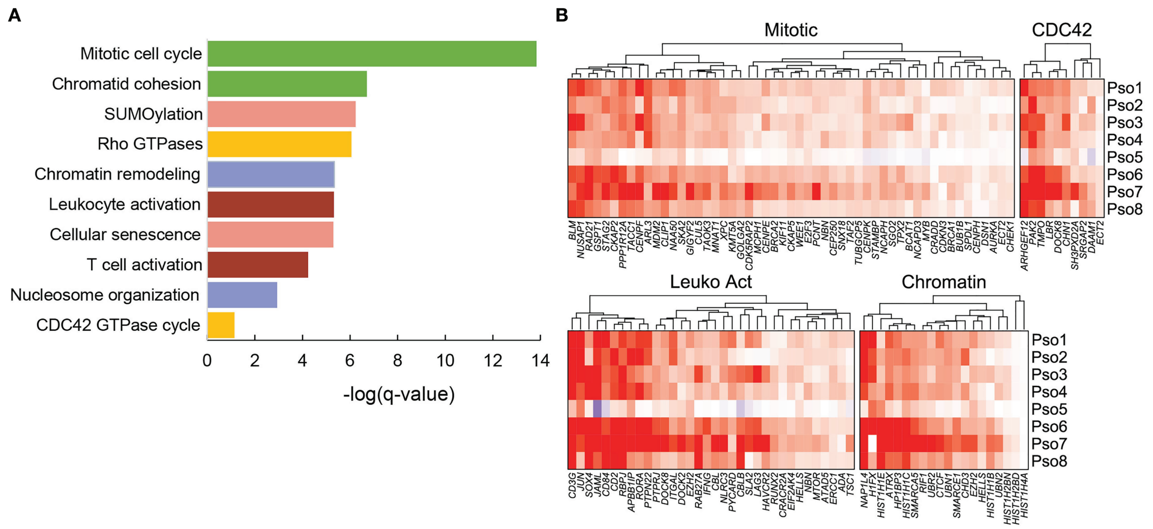


FIGURE 4 | Significant functional associations for the 662 genes significantly elevated in psoriasis samples compared to grouped healthy controls in Trm2. **(A)** Ten example classifications are shown, with functions such as immune cell activation, mitotic cell division, and cytoskeletal reorganization. **(B)** Heatmaps visually represent average log2FC between individual psoriasis samples and normal controls using ComplexHeatmap (18). Heterogeneity is detected between patients, most prominently the dampened amplitude of transcript abnormalities in Patient 5, who was on systemic tacrolimus and mycophenolate at time of biopsy.

in each pathway, activation parallels severity of individual lesions, revealing a potential source of some proportion of clinical heterogeneity.

Psoriasis Single Cells Expressing High Levels of Pathogenic Cytokines Display Elevated T Cell Activation and Cytoskeletal Reorganization Genes

While pervasive elevation of transcripts regulating mitosis or CDC42-centric functional reorganization coincided with induction of pathogenic cytokines in the Trm compartment, we were uncertain whether these programs were related on the single-cell level. In one model, members of these programs might simply be stochastically elevated in any given, pathogenically IL23-polarized single T cell. Alternatively, we hypothesized that some of these transcriptional programs could be shared on the single-cell level, a pattern that could impact approaches to therapeutic targeting. For example, if the single T cells most likely to express Th17 cytokines also showed robust reprogramming of cytoskeleton genes, strategies restraining actin reorganization might impede the mobility and infiltration of the most pathogenic skin-resident T cells.

We therefore calculated the Pearson correlation coefficients for expression of pathogenic cytokines in single Trm2 cells against those of genes in our cytoskeletal and secretory classes, finding striking instances of both positive and negative correlation (**Figure 5**). For example, **Figure 5A** shows positive correlation of the *RORA* transcription factor with *IL17F* expression, as would be expected given its role in Th17 programming (R = 0.3, p = 2.2 x 10^{-16}) (25), as well as for the TCR component *CD3G* (R = 0.29, p = 2.2 x 10^{-16}). Similarly, in **Figure 5B**, the cytoskeletal re-organization genes *PAK2* and *APBB1IP* robustly positively correlate with *IL17F* expression in single skin-resident memory T cells, supporting a model in which more highly pathogenically activated cells are also more motile and capable of tissue infiltration. In sharp contrast, the single cells expressing maximum *IL17F* and those expressing elevated levels of a number of chromatin-modifying transcripts are negatively correlated, for example, a R of -0.30 for *HIST1H1E* (**Figure 5C**). Such instances of mutual exclusivity suggest the presence of a second, abnormal, non-Th17 population within psoriatic Trm, whose influence on disease state is yet undetermined. A comprehensive single-cell correlation table in Trm2 for *IL17F*, *CXCL13*, and *IL26* is available in **Supplementary Table 7**.

DISCUSSION

While a vast landscape of transcriptional abnormalities in immune and stromal cell types characterizes chronic inflammatory skin disease (26, 27), clinical improvement following inhibition of the IL12/23 pathway or blockade of IL17 isoforms validates the central role of psoriatic T cells. Our single-cell profiles of 8 psoriasis samples, along with normal controls, begin to illuminate patient-specific variation of

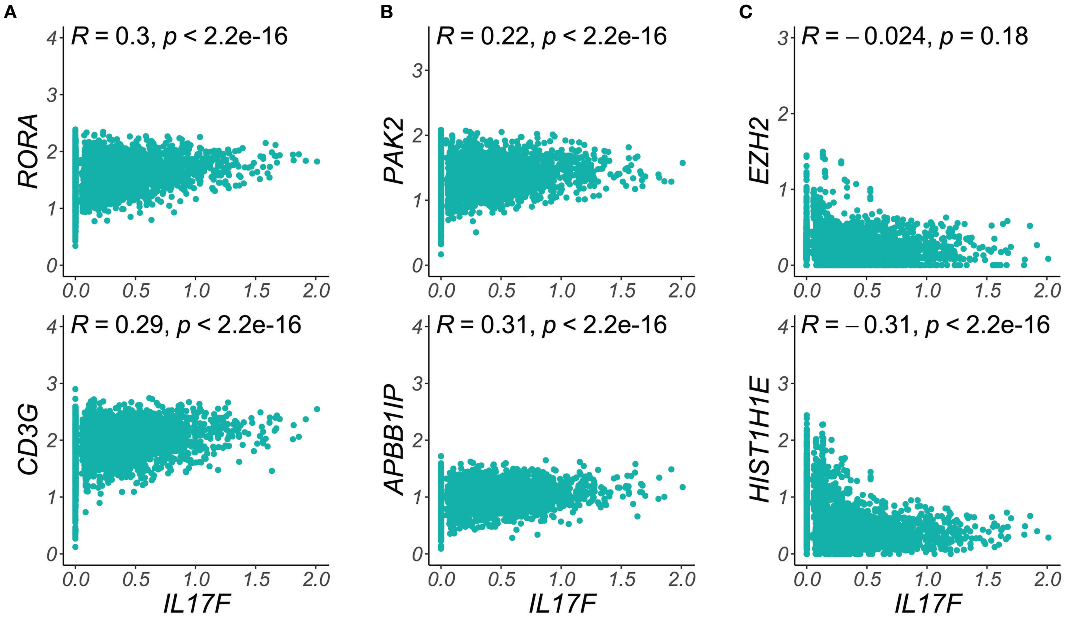


FIGURE 5 | Single-cell correlations and anti-correlations between functional class transcripts and *IL17F* expression. **(A)** T cell activation markers like *RORA* and *CD3G* are elevated in the highest *IL17F* expressing cells, **(B)** Key cytoskeletal reorganization transcripts (*PAK2, APBB1IB*) are most elevated in the single skin-resident T cells expressing maximal psoriatic inflammatory mediators. **(C)** Chromatin remodeling transcripts (*EZH2*, *HIST1H1E*), are elevated in the lowest *IL17F* expression cells, suggesting a distinct, pathologic cell population in psoriatic Trm. Density plots show imputed single cell expression of T cell activation, chromatin remodeling, or cytoskeletal transcripts (*y*-axis) vs. *IL17F* (*x*-axis). Dots represent single Trm2 cells (psoriasis samples).

transcriptional abnormalities in psoriatic Trm (**Figure 2**). Psoriasis DEGs with greater than average patient-level variance, reflected in higher dispersion scores, include the most recognizable Th17 cytokines such as *IL17F* and *IL26*, recently implicated inflammatory psoriatic mediators such as *CXCL13* (13), and genes orchestrating cell division in mitotically active T cells. Such psoriasis DEGs with higher dispersion scores may represent sources of patient-specific phenotypic and clinical variability, such as lesion intensity or anatomic distribution.

Expression of *CXCL13* and *CD84* correlated significantly with lesional severity score in our study, a predicate for further investigation as sources or important associations of disease state. Our data adds to increasing evidence that *CXCL13* represents a particularly Th17-specific abnormality (8) and positively associates with psoriasis severity (13, 28), nominating it as a clinically useful biomarker for cutaneous disease. *CD84* is a known T cell activation marker, genetic variants of which have been associated with response in psoriasis to TNF blockade (29). Interestingly, *IL17F* expression in our series correlated poorly with lesional severity score but was highly elevated in scalp psoriasis, suggesting it might show anatomic specificity in more highly powered studies. This finding comports with an earlier scRNA-seq report that Th17 cytokine expression and overall inflammatory state is surprisingly prominent in healthy scalp cells (27). The key constraint of our study is patient number, limited by the current costs of scRNA-seq. It is very likely additional such correlations will reach significance as these approaches are extended to larger data sets.

Conversely, psoriasis-specific DEGs harboring lower dispersion scores may be more suitable for broader screening to identify psoriasis-like molecular profiles, a feature that may help direct biological treatment for the subset of rashes demonstrating both psoriasiform and spongiotic histopathology (30). Within the set of psoriasis-specific skin-resident DEGs that are overexpressed relative to analogous T cells in atopic dermatitis, examples of such lower variance Th17 biomarkers include the GTPase-activator *CHN1* (0.069) and *PTMS* (0.077).

We also undertook a systematic search of coordinated functional derangements in skin-resident T cells, based on the increased resolution afforded by single-cell transcriptomics. Such groups of pathologic transcriptional alternations may function as quantitative traits, collectively modifying disease phenotype beyond the impact of dysregulated single genes. Applying this method, we detected not only expected elevations in inflammatory signalling and cell division, but also global increases in pathways coordinating CDC42-centric cytoskeletal reorganization and chromatin remodeling. In one sense, broad alterations in these programs are not surprising, given the profound changes in cell polarity and motility that accompany T cell activation. However, this is the first report describing recurrent upregulation of dozens of these transcripts in pathologically inflamed T cells. All 8 patients in our series show abnormalities in these programs (**Figure 4**), whose elevation trends with lesional psoriasis severity scores, supporting a role in the pathogenicity of skin inflammation.

We also show that single T cells expressing the highest levels of psoriatic inflammatory mediators such as *IL17F* are markedly enriched for cytoskeletal remodeling transcripts, suggesting such programs may facilitate tissue infiltration and cytokine secretion. Combination therapeutic approaches targeting both Th17 polarization and cytoskeletal activity may thus synergistically target a common population of particularly pathogenic skin T cells. We also find that certain chromatin remodeling DEGs peak in single T cells distinct from those maximally expressing *IL17F*, indicating these data can also identify additional, abnormally reprogrammed subpopulations within the Trm compartment.

In summary, the analyses presented here describe a suite of quantitative approaches to evaluate high-resolution transcriptional variation between psoriasis patients. The most distinguishing abnormalities are identified in skin-resident T cells, and even our limited test set identifies credible associations between specific genes and lesion phenotype. Greater numbers of scRNA-seq datasets are now becoming publicly accessible. Systematic identification of such instances of inter-individual molecular heterogeneity will make it possible to test clinically predictive associations for both single genes and aggregate molecular disease signatures.

DATA AVAILABILITY STATEMENT

Sequencing BAM files are deposited at the European Genome-Phenome Archive (EGA) under accession number EGAS 00001005271. The processed Seurat object with cell identities is available at Zenodo (https://zenodo.org/record/6529821#.Ynh_aoxBwuW).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of California, San Francisco Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YL, JC, EP, and RC designed the study. EK, JC, and RC supervised sample collection and processing. YL, CC, AH, and AS performed sample preparation and analysis. HW, MT, CC, YS, and EP performed computational analysis. JN performed histopathology analysis. YL, HW, EP, JC, and RC wrote the original manuscript with contributions from CC, JN, and AH. All authors contributed to the article and approved the submitted version.

FUNDING

This study received funding from Sun Pharmaceutical Industries. The funder was not involved in the study design, collection, analysis, and interpretation of data, the writing of this article or the decision to submit it for publication. JC and RC are funded in part by grants from the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health K08AR067243, the National Psoriasis Foundation, and a LEO foundation grant.

ACKNOWLEDGMENTS

We acknowledge Rachel Sevey for essential assistance with graphic representation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.842651/full#supplementary-material

Supplementary Table 1 | Clinical Characteristics of Psoriasis Patients.

Supplementary Table 2 | Cell Type Marker Transcripts (Seurat-derived).

Supplementary Table 3 | Trm2 Pseudo-bulk DEGs Normal Controls vs. Psoriasis.

Supplementary Table 4 | Deduced Cell Number Expression for each Gene.

Supplementary Table 5 | Correlation of Gene Expressing Cells with Lesional Psoriasis Severity Scores.

Supplementary Table 6 | Metascape Functional Analysis of Trm2 Psoriasis DEGs.

Supplementary Table 7 | Single-cell gene correlations (non-imputed) in Trm2 for *IL17F*, *CXCL13*, and *IL26*.

Supplementary Data Sheet 1 | uMAP representation of donors, showing high representation for each sample in key immune cell classes.

Supplementary Data Sheet 2 | Box plots of selected genes from Trm2 DEG enriched functional classes on a per-sample basis.

REFERENCES

1. Bugaut H, Aractingi S. Major Role of the IL17/23 Axis in Psoriasis Supports the Development of New Targeted Therapies. *Front Immunol* (2021) 12:621956. doi: 10.3389/fimmu.2021.621956
2. Itoh T, Hatano R, Komiya E, Otsuka H, Narita Y, Aune TM, et al. Biological Effects of IL-26 on T Cell-Mediated Skin Inflammation, Including Psoriasis. *J Invest Dermatol* (2019) 139(4):878–89. doi: 10.1016/j.jid.2018.09.037
3. Jordan CT, Cao L, Roberson EDO, Duan S, Helms CA, Nair RP, et al. Rare and Common Variants in CARD14, Encoding an Epidermal Regulator of NF-kappaB, in Psoriasis. *Am J Hum Genet* (2012) 90(5):796–808. doi: 10.1016/j.ajhg.2012.03.013
4. Li B, Tsoi LC, Swindell WR, Gudjonsson JE, Tejasvi T, Johnston A, et al. Transcriptome Analysis of Psoriasis in a Large Case-Control Sample: RNA-seq Provides Insights Into Disease Mechanisms. *J Invest Dermatol* (2014) 134 (7):1828–38. doi: 10.1038/jid.2014.28
5. Keermann M, Kõks S, Reimann E, Prans E, Abram K, Kingo K. Transcriptional Landscape of Psoriasis Identifies the Involvement of IL36 and IL36RN. *BMC Genomics* (2015) 16(1):322. doi: 10.1186/s12864-015-1508-2
6. Visvanathan S, Baum P, Vinisko R, Schmid R, Flack M, Lalovic B, et al. Psoriatic Skin Molecular and Histopathologic Profiles After Treatment With Risankizumab Versus Ustekinumab. *J Allergy Clin Immunol* (2019) 143 (6):2158–69. doi: 10.1016/j.jaci.2018.11.042
7. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, et al. Simultaneous Epitope and Transcriptome Measurement in Single Cells. *Nat Methods* (2017), 14 (9):865–8. doi: 10.1038/nmeth.4380.
8. Liu Y, Wang H, Taylor MA, Cook C, Martinez-Berdeja A, North JP, et al. Classification of Human Chronic Inflammatory Skin Disease Based on Immune Single-Cell Profiling. *Sci Immunol*.
9. Crowell HL, Soneson C, Germain P-L, Calini D, Collin L, Raposo C, et al. Muscat Detects Subpopulation-Specific State Transitions From Multi-Sample Multi-Condition Single-Cell Transcriptomics Data. *Nat Commun* (2020) 11 (1):6077. doi: 10.1038/s41467-020-19894-4
10. *edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data* (2022). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2796818/.
11. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets. *Nat Commun* (2019) 10(1):1523. doi: 10.1038/s41467-019-09234-6
12. *Metascape* (2021). Available at: https://metascape.org/gp/index.html#/main/step1.
13. Liu J, Chang H-W, Huang Z-M, Nakamura M, Sekhon S, Ahn R, et al. Single-Cell RNA Sequencing of Psoriatic Skin Identifies Pathogenic Tc17 Cell Subsets and Reveals Distinctions Between CD8+ T Cells in Autoimmunity and Cancer. *J Allergy Clin Immunol* (2021) 147(6):2370–80. doi: 10.1016/j.jaci.2020.11.028
14. *estimateTagwiseDisp Function - Rdocumentation* (2022). Available at: https://www.rdocumentation.org/packages/edgeR/versions/3.14.0/topics/estimateTagwiseDisp.
15. Mahil SK, Catapano M, Di Meglio P, Dand N, Ahlfors H, Carr IM, et al. An Analysis of IL-36 Signature Genes and Individuals With IL1RL2 Knockout Mutations Validates IL-36 as a Psoriasis Therapeutic Target. *Sci Transl Med* (2017) 9(411):eaan2514. doi: 10.1126/scitranslmed.aan2514
16. Yap JY, Moens L, Lin M-W, Kane A, Kelleher A, Toong C, et al. Intrinsic Defects in B Cell Development and Differentiation, T Cell Exhaustion and Altered Unconventional T Cell Generation Characterize Human Adenosine Deaminase Type 2 Deficiency. *J Clin Immunol* (2021) 41(8):1915–35. doi: 10.1007/s10875-021-01141-0
17. Cosmi L, De Palma R, Santarlasci V, Maggi L, Capone M, Frosali F, et al. Human Interleukin 17-Producing Cells Originate From a CD161+CD4+ T Cell Precursor. *J Exp Med* (2008) 205(8):1903–16. doi: 10.1084/jem.20080397
18. Gu Z, Eils R, Schlesner M. Complex Heatmaps Reveal Patterns and Correlations in Multidimensional Genomic Data. *Bioinf* (2016) 32 (18):2847–9. doi: 10.1093/bioinformatics/btw313
19. Ghanem NZ, Matter ML, Ramos JW. Regulation of Leukaemia Associated Rho Gef (Larg/Arhgef12). *Small GTPases* (2021) 1–9. doi: 10.1080/21541248.2021.1951590
20. Wang Y, Zeng C, Li J, Zhou Z, Ju X, Xia S, et alPak2 Haploinsufficiency Results in Synaptic Cytoskeleton Impairment and Autism-Related Behavior. (2022).
21. Biggs CM, Keles S, Chatila TA. DOCK8 Deficiency: Insights Into Pathophysiology, Clinical Features and Management. *Clin Immunol* (2017) 181:75–82. doi: 10.1016/j.clim.2017.06.003
22. Ichihara-Tanaka K, Kadomatsu K, Kishida S. Temporally and Spatially Regulated Expression of the Linker Histone H1fx During Mouse Development. *J Histochem Cytochem* (2017) 65(9):513–30. doi: 10.1369/0022155417723914
23. Rodriguez P, Munroe D, Prawitt D, Chu LL, Bric E, Kim J, et al. Functional Characterization of Human Nucleosome Assembly Protein-2 (NAP1L4) Suggests a Role as a Histone Chaperone. *Genomics* (1997) 44(3):253–65. doi: 10.1006/geno.1997.4868
24. Ding Y, Li Y, Zhao Z, Cliff Zhang Q, Liu F. The Chromatin-Remodeling Enzyme Smarca5 Regulates Erythrocyte Aggregation *via* Keap1-Nrf2 Signaling. *Elife* (2021) 10:e72557. doi: 10.1101/2021.09.08.459391

25. Yang XO, Pappu B, Nurieva R, Akimzhanov A, Kang HS, Chung Y, et al. TH17 Lineage Differentiation Is Programmed by Orphan Nuclear Receptors Rorα and Rorγ. *Immun* (2008) 28(1):29–39. doi: 10.1016/j.immuni.2007.11.016
26. Reynolds G, Vegh P, Fletcher J, Poyner EFM, Stephenson E, Goh I, et al. Developmental Cell Programs Are Co-Opted in Inflammatory Skin Disease. *Science* (2021) 371(6527):eaba6500. doi: 10.1126/science.aba6500
27. Cheng JB, Sedgewick AJ, Finnegan AI, Harirchian P, Lee J, Kwon S, et al. Transcriptional Programming of Normal and Inflamed Human Epidermis at Single-Cell Resolution. *Cell Rep* (2018) 25(4):871–83. doi: 10.1016/j.celrep.2018.09.006
28. Liu W, Zhou X, Wang A, Ma J, Bai Y. Increased Peripheral Helper T Cells Type 17 Subset Correlates With the Severity of Psoriasis Vulgaris. *Immunol Lett* (2021) 229:48–54. doi: 10.1016/j.imlet.2020.11.005
29. van den Reek JMPA, Coenen MJH, van de L'Isle Arias M, Zweegers J, Rodijk-Olthuis D, Schalkwijk J, et al. Polymorphisms in CD84, IL12B and TNFAIP3 Are Associated With Response to Biologics in Patients With Psoriasis. *Br J Dermatol* (2017) 176(5):1288–96. doi: 10.1111/bjd.15005
30. Cohen JN, Bowman S, Laszik ZG, North JP. Clinicopathologic Overlap of Psoriasis, Eczema, and Psoriasiform Dermatoses: A Retrospective Study of T Helper Type 2 and 17 Subsets, Interleukin 36, and β-Defensin 2 in Spongiotic Psoriasiform Dermatitis, Sebopsoriasis, and Tumor Necrosis Factor α Inhibitor-Associated Dermatitis. *J Am Acad Dermatol* (2020) 82(2):430–9. doi: 10.1016/j.jaad.2019.08.023

Conflict of Interest: JC and RC receive support (research grants to their institution) from LEO Pharmaceuticals, Sanofi, and Pfizer.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership