frontiers | Research Topics

Genetics of inflammatory and immune diseases

Edited by
Yonghu Sun, Yunqing Ren and Xianyong Yin

Published in
Frontiers in Genetics

FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-4334-4
DOI 10.3389/978-2-8325-4334-4

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

Genetics of inflammatory and immune diseases

Topic editors

Yonghu Sun — Shandong Provincial Hospital of Dermatology, China

Yunqing Ren — Zhejiang University, China

Xianyong Yin — University of Michigan, Ann Arbor, United States

Citation

Sun, Y., Ren, Y., Yin, X., eds. (2024). *Genetics of inflammatory and immune diseases*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4334-4

Table of contents

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 08 January 2024
DOI 10.3389/fgene.2023.1355794

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Jordi Pérez-Tur,
Spanish National Research Council (CSIC), Spain

*CORRESPONDENCE
Yonghu Sun,
suohandong@126.com

RECEIVED 14 December 2023
ACCEPTED 27 December 2023
PUBLISHED 08 January 2024

CITATION
Wang T, Ren Y, Yin X and Sun Y (2024), Editorial: Genetics of inflammatory and immune diseases. *Front. Genet.* 14:1355794. doi: 10.3389/fgene.2023.1355794

COPYRIGHT
© 2024 Wang, Ren, Yin and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: Genetics of inflammatory and immune diseases

Tianyu Wang[1,2], Yunqing Ren[3], Xianyong Yin[4] and Yonghu Sun[1,2]*

[1]Hospital for Skin Diseases, Shandong First Medical University, Jinan, Shandong, China, [2]Shandong Provincial Institute of Dermatology and Venereology, Shandong Academy of Medical Sciences, Jinan, Shandong, China, [3]Department of Dermatology, Children's Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, China, [4]Department of Biostatistics, University of Michigan, Ann Arbor, MI, United States

KEYWORDS

inflammatory and immune diseases, susceptibility, genome-wide association study, high-throughput sequencing, integrative study, mechanism

Editorial on the Research Topic
Genetics of inflammatory and immune diseases

1 Introduction

Inflammatory and immune diseases encompass a spectrum of conditions, including inflammatory bowel disease, systemic lupus erythematosus, rheumatoid arthritis, psoriasis, and atopic dermatitis (Hodson, 2016; Sroka-Tomaszewska and Magdalena, 2021; Kiriakidou and Ching, 2020). These disorders are characterized by presumed pathomechanisms involving inflammatory and immune systems, frequently exhibiting a significant genetic predisposition. Over the past few decades, genetic studies, particularly genome-wide association studies (GWAS), have identified hundreds of genetic variants robustly associated with these diseases (Klein et al., 2005). The rapid accumulation of GWAS data has broadened opportunities for biologists to discover new disease-associated variants and gain insights into the mechanisms underlying complex human diseases (Wang et al., 2016; Sun et al., 2019; Wang et al., 2021). Although GWAS is a powerful tool, it encounters challenges in precisely pinpointing candidate disease risk genes. In response to these challenges, several post-GWAS methods have emerged, including Transcriptome-Wide Association Studies (TWAS) (Gusev et al., 2016) and Mendelian Randomization (MR) (Bowden and Holmes, 2019). These methods serve as potent tools for identifying candidate disease risk genes, providing benefits such as enhanced statistical power, improved interpretability, and reduced computational costs. Importantly, the exploration of genetic mechanisms (Stikker et al., 2023; Sun et al., 2016), experimental integration research (Xia et al., 2022), multi-omics studies (Husin et al., 2017), and advancements in sequencing technology (Mi et al., 2022) has further contributed to new biological insights into the mechanisms of inflammation and immune disorders. These findings have the potential to identify drug targets of interest.

2 Overview of contributions

The identification of disease-causing variants is crucial for enhancing our understanding of disease pathogenesis and expediting the discovery of diagnostic biomarkers for inflammatory and immune disorders. Fang et al. reported a case study revealing a novel heterozygous variant (c.934G>T, p.Glu312Ter) in the MSN gene associated with X-linked moesin-associated immunodeficiency (X-MAID), identified through whole exome sequencing (WES) and trio analysis. This study broadened the spectrum of MSN mutations linked to immunodeficiency. Li et al. conducted an integrated bioinformatics analysis of 30 patients with systemic lupus erythematosus (SLE) to identify key genes in $CD4^+$ T cells, uncovering six novel biomarkers that could contribute to the diagnosis and treatment of SLE. Fu et al. identified immune biomarkers associated with basement membranes in idiopathic pulmonary fibrosis (IPF) through integrated bioinformatics and Pan-cancer analysis. The study pinpointed immune-related hub genes that could serve as potential targets for intervention in various diseases, including IPF and cancer. Additionally, Tang et al. utilized genome sequencing analysis (WGS), meta-analyses, and replication analyses to discover a rare gain-of-function frameshift variant in SIRPB1 associated with Crohn's disease (CD) in Han Chinese patients. The study also investigated the functional mechanism of SIRPB1 and its downstream inflammatory pathways in CD.

The integration of experimental studies with multi-omics studies holds immense potential for unraveling the intricate mechanisms underlying diseases. Yin et al. integrated experimental and transcriptomic approaches, revealing the involvement of interleukin enhancer binding factor 2 (ILF2) and KLHDC7B-DT in the hyperproliferation of keratinocytes and skin inflammation in psoriasis. Notably, ILF2 operates in a KLHDC7B-DT-dependent manner. Similarly, Yan et al. integrated experimental and transcriptomic data to elucidate the potential involvement of serum extracellular vesicles (EVs) in the activation of keratinocytes through loaded miRNAs in psoriasis. Specific miRNAs, such as miR-1305 and miR-6785-5p within serum EVs, were identified as potential contributors to psoriasis. Duan et al. adopted an integrative approach, combining experimental and bioinformatics analyses of 38 samples, to reveal a potential regulatory mechanism of m6A modification in the immune microenvironment of osteoarthritis (OA). This not only provides insights for OA treatment but also addresses a research gap in this field. Furthermore, Ribeiro et al. employed an integrated approach combining experimental and bioinformatics analyses to reveal the global haplotype-specific 3-dimensional chromatin looping architecture. This architecture significantly influences local allelic BLK and FAM167A gene expression, offering mechanistic details on how regional variants controlling the BLK promoter may impact disease risk.

The current study explores the genetic associations and pathogenesis of inflammatory and immune diseases through the application of GWAS and other analytical approaches. Elghzaly et al. performed a GWAS in an admixed Egyptian population, revealing novel novel genetic associations and providing insights into the pathogenesis of SLE. Their work contributes to an enhanced understanding of the genetic factors underlying this autoimmune condition. In a related study, Zhang et al. employed multifactor dimensionality reduction (MDR) analysis and meta-analysis techniques to elucidate the interaction between ERAP1 and IFIH1 in influencing the development of psoriasis. This approach provides valuable insights into the complex interplay of genetic factors contributing to the pathophysiology of psoriasis. Bui et al. conducted an investigation into the contributions of both HLA and non-HLA factors to clinical phenotypes within the Newfoundland psoriasis cohort. Utilizing a partitioned 88-loci psoriasis Genetic Risk Score (GRS), they successfully clarified the intricate relationship between HLA and non-HLA components of the GRS with crucial clinical features of psoriasis.

Addressing the challenging task of distinguishing causal signals from mere associations in GWAS poses a substantial challenge. To address this challenge, various methods including fine mapping, MR, and TWAS have emerged, offering avenues to translate GWAS findings into a functional understanding of associated traits. Lee et al. employed meticulous fine mapping to identify the MHC locus variant linked to late-onset asthma across diverse race/ethnicity-stratified populations. This precision enhances our understanding of the genetic basis of asthma and holds promise for the development of precise, targeted treatments tailored to specific demographic groups. Zhao et al. employed a different approach by utilizing circulating cytokines and conducting a two-sample MR analysis to identify causal cytokines associated with the risk of psoriasis vulgaris. This unveiled potential therapeutic targets and shed light on the underlying genetic factors contributing to the development of this dermatological condition. Yang et al. investigated genetically regulated 25OHD concentrations through a meta-analysis and MR study, establishing a causal link between genetically regulated 25OHD concentrations and the risk of chronic obstructive pulmonary disease (COPD). This insight provides a foundation for understanding the genetic underpinnings of COPD and may inform preventive strategies. Zhu et al. conducted a comprehensive strategy involving TWAS and multi-tissue interaction analyses. Their work unveiled candidate susceptibility genes related to respiratory infectious diseases, offering valuable insights at the genetic level and potentially informing future interventions and therapies.

Numerous studies highlighted in this Research Topic have explored diverse phenotypic anchoring strategies to establish genotype-phenotype associations in the realm of inflammatory and immune diseases. In a comprehensive genotype-phenotype association study, Chen et al. employed a multifaceted approach, incorporating clinical characteristics, *in silico* analysis, and intervention. Their investigation led to the identification of two novel compound heterozygous mutations in TTC7A, a groundbreaking discovery marking the first instance of these mutations observed in mainland China. These mutations were associated with neonatal-onset Inflammatory Bowel Disease-Combined Immunodeficiency (IBD-CID) and linked to a poor prognosis. Similarly, Xu et al. delved into the realm of congenital pseudarthrosis of the tibia (CPT) combined with neurofibromatosis type 1, employing a novel NF1 mutation. Their study utilized a combination of clinical characteristics and *in silico* analysis to unravel the intricacies of this distinctive genotype-phenotype association. This research adds valuable insights into the understanding of the complex interplay between genetic

mutations and clinical manifestations in the realm of inflammatory and immune diseases.

Mini-reviews summarizing reported genetic variants, especially in the realm of inflammatory and immune diseases, serve as a valuable approach for hypothesis generation and deepening our understanding of underlying molecular mechanisms. Wang et al. conducted a review focusing on the roles of the AIM2 gene and AIM2 inflammasome in the pathogenesis and treatment of psoriasis. This mini-review aims to provide not only a comprehensive overview but also new insights into the functions of the AIM2 gene and AIM2 inflammasome for psoriasis.

3 Conclusion

Collectively, these articles unveil the genetic foundations of inflammatory and immune diseases. They not only elucidate the mechanisms underlying genetic associations but also uncover new therapeutic targets through the comprehensive application of GWAS, both analytically and experimentally. Post-GWAS, these studies integrate analytics and experiments to deepen our understanding of genetic mechanisms, leveraging advanced sequencing technologies. In doing so, these contributions stand as valuable knowledge resources for future medical research and clinical applications. By offering insights into the intricate genetic landscape of inflammatory and immune diseases, these studies contribute significantly to the ongoing advancement of medical knowledge, fostering the potential for innovative treatments and improved clinical outcomes.

Author contributions

TW: Investigation, Writing–original draft, Writing–review and editing. YR: Writing–original draft, Writing–review and editing. XY: Writing–original draft, Writing–review and editing. YS: Writing–original draft, Writing–review and editing.

Acknowledgments

We thank the authors who contributed their work to this Research Topic "*Genetics of Inflammatory and Immune Diseases*."

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.

References

Bowden, J., and Holmes, M. V. (2019). Meta-analysis and Mendelian randomization: a review. *Res. Synth. Methods.* 10 (4), 486–496. doi:10.1002/jrsm.1346

Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B. W. J. H., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* 48 (3), 245–252. doi:10.1038/ng.3506

Hodson, R. (2016). Inflammatory bowel disease. *Nature* 540 (7634), S97. doi:10.1038/540S97a

Husin, Y., Seldin, M., and Lusis, A. (2017). Multi-omics approaches to disease. *Genome. Biol.* 18 (1), 83. doi:10.1186/s13059-017-1215-1

Kiriakidou, M., and Ching, C. L. (2020). Systemic lupus erythematosus. *Ann. Intern. Med.* 172 (11), ITC81–ITC96. doi:10.7326/AITC202006020

Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J. Y., Sackler, R. S., Haynes, C., et al. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science* 308 (5720), 385–389. doi:10.1126/science.1109557

Mi, Z., Wang, Z., Xue, X., Liu, T., Wang, C., Sun, L., et al. (2022). The immune-suppressive landscape in lepromatous leprosy revealed by single-cell RNA sequencing. *Cell. Discov.* 8 (1), 2. doi:10.1038/s41421-021-00353-3

Sroka-Tomaszewska, J., and Magdalena, T. M. (2021). Molecular mechanisms of atopic dermatitis pathogenesis. *Int. J. Mol. Sci.* 22 (8), 4130. doi:10.3390/ijms22084130

Stikker, B. J., Hendriks, R. W., and Stadhouders, R. (2023). Decoding the genetic and epigenetic basis of asthma. *Allergy* 78 (4), 940–956. doi:10.1111/all.15666

Sun, Y., Irwanto, A., Toyo-Oka, L., Hong, M., Liu, H., Andiappan, A. K., et al. (2016). Fine-mapping analysis revealed complex pleiotropic effect and tissue-specific regulatory mechanism of TNFSF15 in primary biliary cholangitis, Crohn's disease, and leprosy. *Sci. Rep.* 6, 31429. doi:10.1038/srep31429

Sun, Y., Liu, H., Yang, B., Wang, C., Foo, J. N., Bao, F., et al. (2019). Investigation of the predisposing factor of pemphigus and its clinical subtype through a genome-wide association and next-generation sequence analysis. *J. Eur. Acad. Dermatol. Venereol.* 33 (2), 410–415. doi:10.1111/jdv.15227

Wang, Y., Zhang, Y., Lin, Z., Zhang, H., Wang, T., Cao, Y., et al. (2021). Identification of 38 novel loci for systemic lupus erythematosus and genetic heterogeneity between ancestral groups. *Nat. Commun.* 12 (1), 772. doi:10.1038/s41467-021-21049-y

Wang, Z., Sun, Y., Fu, X., Yu, G., Wang, C., Bao, F., et al. (2016). A large-scale genome-wide association and meta-analysis identified four novel susceptibility loci for leprosy. *Nat. Commun.* 15 (7), 13760. doi:10.1038/ncomms13760

Xia, Q., Lyu, C., Li, B., Pang, B., Guo, X., Ren, H., et al. (2022). Candidate drugs screening for behcet's disease based on bioinformatics analysis and mouse experiments. *Front. Immunol.* 13, 895869. doi:10.3389/fimmu.2022.895869

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 02 May 2022
doi: 10.3389/fgene.2022.890624

ILF2 Contributes to Hyperproliferation of Keratinocytes and Skin Inflammation in a KLHDC7B-DT-Dependent Manner in Psoriasis

Xiran Yin[1,2], Zhenxian Yang[1,2], Mingsheng Zhu[3], Cheng Chen[1,2], Shan Huang[1,2], Xueqing Li[1,2], Hua Zhong[1,2], He Wen[1,2], Qing Sun[1,2], Xiaojing Yu[1,2] and Jianjun Yan[1,2]**

[1]Department of Dermatology, Qilu Hospital, Shandong University, Jinan, China, [2]Laboratory of Basic Medical Science, Qilu Hospital, Shandong University, Jinan, China, [3]Department of Hand and Foot surgery, Shandong Provincial Hospital, Jinan, China

OPEN ACCESS

Edited by:
Yonghu Sun,
Shandong Provincial Hospital of Dermatology, China

Reviewed by:
Haijing Wu,
Central South University, China
Erle Dang,
Fourth Military Medical University, China

***Correspondence:**
Xiaojing Yu
yuxiaojing96@163.com
Jianjun Yan
jianjun19870103@163.com

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: 06 March 2022
Accepted: 18 April 2022
Published: 02 May 2022

Citation:
Yin X, Yang Z, Zhu M, Chen C, Huang S, Li X, Zhong H, Wen H, Sun Q, Yu X and Yan J (2022) ILF2 Contributes to Hyperproliferation of Keratinocytes and Skin Inflammation in a KLHDC7B-DT-Dependent Manner in Psoriasis. Front. Genet. 13:890624. doi: 10.3389/fgene.2022.890624

Background: The extensive involvement of interleukin enhancer binding factor 2 (ILF2) in RNA stability and the inflammatory response is well documented. Aberrant long noncoding RNA (lncRNA) expression contributes to the pathogenesis of psoriasis. However, little is known about the role of ILF2 in psoriasis.

Objective: To investigate the role of ILF2 and KLHDC7B-DT in psoriasis.

Methods: LncRNA expression in psoriatic tissues was measured by lncRNA microarray and qRT–PCR. Normal human epidermal keratinocytes (NHEKs), HaCaT cells, and Ker-CT cells stimulated with M5 (IL-17A, IL-22, IL-1α, oncostatin M, and TNF-α) were used to establish a psoriasis model *in vitro*. Fluorescence *in situ* hybridization was used to detect the distribution of KLHDC7B-DT and ILF2 in keratinocytes. The proliferative effects of KLHDC7B-DT and ILF2 on keratinocytes were demonstrated by EdU assay and flow cytometry. ELISA was used to detect the secretion levels of cytokines. RNA pull-down and RNA immunoprecipitation (RIP) were used to detect the direct binding of KLHDC7B-DT with ILF2. Western blotting was used to detect the proteins related to STAT3/JNK signalling pathways.

Results: ILF2 and KLHDC7B-DT were significantly overexpressed in psoriatic tissues and M5-induced keratinocytes. KLHDC7B-DT promoted the proliferation of keratinocytes and induced the secretion of IL-6 and IL-8. KLHDC7B-DT could directly bind to ILF2 and activate the STAT3 and JNK signalling pathways. KLHDC7B-DT expression was regulated by ILF2. M5-induced proliferation and inflammatory cytokine secretion in keratinocytes was inhibited after ILF2 knockdown. Furthermore, we found that ILF2 promoted keratinocyte proliferation and the inflammatory response in a KLHDC7B-DT-dependent manner.

Conclusions: ILF2 and KLHDC7B-DT are involved in the hyperproliferation of keratinocytes and skin inflammation in psoriasis. In addition, ILF2 functions in a KLHDC7B-DT-dependent manner.

Keywords: ILF2, KLHDC7B-DT, psoriasis, proliferation, inflammation

INTRODUCTION

Psoriasis is a common immune-mediated chronic inflammatory skin disease characterized by red, scaly skin plaques (Boehncke and Schön, 2015; Srivastava et al., 2017). Recent studies have shown that immune, genetic and environmental factors are all responsible for the pathogenesis of psoriasis (Nickoloff et al., 2007a; Lin et al., 2011). Accumulating studies have suggested that psoriasis is mainly caused by abnormal crosstalk between keratinocytes and immune cells. Various cytokines secreted by immune cells activate keratinocytes and lead to keratinocyte secretion of proinflammatory cytokines, thereby sustaining the inflammatory response (Albanesi et al., 2005; Albanesi et al., 2007).

Long noncoding RNA (lncRNA) is a noncoding RNA with a length of more than 200 nucleotides and limited protein-encoding abilities (Fatica and Bozzoni, 2014; Iyer et al., 2015). Recent studies have confirmed that some inflammatory diseases, such as psoriasis, rheumatoid arthritis (RA) (Wu et al., 2015), Crohn's disease (CD) and ulcerative colitis (UC) (Mirza et al., 2015), and multiple sclerosis (MS) (Santoro et al., 2016), show abnormal lncRNA expression. Psoriasis susceptibility-related RNA gene induced by stress (PRINS) was initially identified as a biomarker of genetic susceptibility to psoriasis (Szegedi et al., 2010). Our previous study confirmed that lncRNA MSX2P1 was involved in the pathogenesis of psoriasis by directly binding to miR-6731-5p and activating S100A7 (Qiao et al., 2018). These findings demonstrate that differentially expressed lncRNAs contribute to the pathogenesis of psoriasis.

In this study, we performed a lncRNA microarray and found 2,194 differentially expressed lncRNAs (1,123 upregulated and 1,071 downregulated; fold change ≥2 and $p < 0.05$). The microarray results showed that ENST00000609178, also named KLHDC7B-DT (NCBI Gene ID: 105373098), was upregulated (fold change of 17.05) in psoriatic lesions compared to normal tissues. Previous reports have shown that KLHDC7B-DT activates STAT3 signalling by inducing IL-6 secretion (Li et al., 2021). STAT3 signalling and IL-6 are closely related to the pathogenesis of psoriasis, and we hypothesized that KLHDC7B-DT might participate in regulating the inflammatory response and be involved in the pathogenesis of psoriasis.

ILF2, also known as nuclear factor 45 (NF45), is crucial for cell growth and the inflammatory response (Du et al., 2019). Recent studies have shown that ILF2 can promote cell proliferation and the inflammatory response (Jia et al., 2020; Zhang et al., 2021). ILF2 forms a complex with the 90 kDa interleukin enhancer binding factor 3 (NF90, ILF3) (Zhao et al., 2005; Kuwano et al., 2008). A previous study confirmed that lncRNAs could promote cell proliferation through association with the NF45/NF90 complex (Huang et al., 2020). In the present study, we found that both ILF2 and KLHDC7B-DT were overexpressed in psoriatic lesions as well as M5-treated keratinocytes compared with controls. KLHDC7B-DT could directly bind with ILF2, and its expression was regulated by ILF2. As KLHDC7B-DT and ILF2 were significantly overexpressed in psoriatic lesions and closely related to inflammatory responses, KLHDC7B-DT and ILF2 were selected as potential candidates for this study.

Here, we explored the role of ILF2 and KLHDC7B-DT in the pathogenesis of psoriasis for the first time. We verified the effects of ILF2 and KLHDC7B-DT on the proliferation of keratinocytes and the secretion of IL-6 and IL-8. In addition, we showed that KLHDC7B-DT promoted the proliferation of keratinocytes and induced the secretion of IL-6 and IL-8 by activating the STAT3 and JNK signalling pathways. Furthermore, we found that ILF2 functions in a KLHDC7B-DT-dependent manner in psoriasis. Our findings provide a potential therapeutic strategy for psoriasis.

MATERIALS AND METHODS

Patient and Tissue Sample Collection

Ten samples were collected from patients with vulgaris psoriasis (6 males and 4 females; aged 24–49 years) from Qilu Hospital of Shandong University. The patients had not received systemic therapy or topical treatment within 3 months. In addition, 10 normal tissues were collected from healthy volunteers (6 males and 4 females; aged 22–40 years). The healthy subjects had no family history of psoriasis or any other autoimmune diseases. We separated the epidermis from psoriatic lesions and healthy control tissues for further experiments. This study was approved by the Ethics Committee of Shandong University, Qilu Hospital (Jinan, China), and all patients provided written informed consent.

LncRNA Microarray

The microarray was performed by Shanghai shbio Biotechnology Co., Ltd. The results were detected by an Agilent microarray scanner and analyzed by Agilent feature extraction software.

Cell Isolation and Culture

Purified normal human epidermal keratinocytes (NHEKs) were obtained after 2–3 generations. HaCaT cells (Procell Life Science & Technology Co., Ltd.) were cultured in Dulbecco's modified Eagle's medium (DMEM) (Gibco, United States) containing 10% foetal bovine serum (FBS) (Sangon Biotech, China), 100 μg/ml streptomycin and 100 U/ml penicillin. Ker-CT (CRL4048™) from American Type Culture Collection (ATCC) was cultured in KGMGold™ BulletKit™ (Lonza 00192060). All cells were incubated in a humidified chamber at 37°C with 5% CO_2.

RNA Fluorescence *In situ* Hybridization

KLHDC7B-DT probes (5′-CY3-TAACGCTCTTTCAGTCAGGTGTTCCCC, GenePharma) were used for hybridization. DAPI (Servicebio, G1012) was used for nuclear staining.

Construction of the Psoriatic Cell Model

When cell confluence reached approximately 60–70%, cells were starved in serum-free DMEM for 12 h. Then, M5 (10 ng/ml final concentration; PeproTech), a cocktail of cytokines, was used to

TABLE 1 | Oligonucleotide sequences.

siRNA	Sense (5′→3′)	Antisense (5′→3′)
siKLHDC7B-DT	GCCCAGUCAUUAUCACAUATT	UAUGUGAUAAUGACUGGGCTT
siILF2	GGACAUUUGAAGUGCAAAUTT	AUUUGCACUUCAAAUGUCCTT
siNC	UUCUCCGAACGUGUCACGUTT	ACGUGACACGUUCGGAGAATT

TABLE 2 | Sequences of primers for qRT–PCR.

RNA	Forward (5′→3′)	Reverse (5′→3′)
LncRNA KLHDC7B-DT	GTTGCTAGTCCTCCGCTTCGC	GCTGGCTTGCCACAGGTTATG
ILF2	CCTTAGCAGCCATCCGACAT	TTAGGGCCAAAGGCTGTCTG
GAPDH	GCACCGTCAAGGCTGAGAAC	TGGTGAAGACGCCAGTGGA
U6	CAGCACATATACTAAAATTGGAACG	ACGAATTTGCGTGTCATCC

treat NHEKs, HaCaT cells, and Ker-CT cells in serum-free DMEM to induce psoriatic inflammation-like conditions for another 24 h. Controls were left untreated.

Cell Transfection

For gene knockdown, small interfering RNAs (siRNAs) (GenePharma) of KLHDC7B-DT and ILF2 were transferred into HaCaT and Ker-CT cells using Lipofectamine 2000 transfection reagent (Invitrogen, Carlsbad, United States). For gene overexpression, HaCaT and Ker-CT cells were transfected with the eukaryotic expression vector pcDNA3.1 (GenePharma, **Supplementary Figure S1**) with jet PRIME transfection reagent (Polyplus, United States). The oligonucleotide sequences are shown in **Table 1**.

qRT–PCR

qRT–PCR was performed as described in our previous experiments (Yan et al., 2019). The relative gene expression was normalized to GAPDH or U6 and calculated using the 2−ΔΔCt method. The primer sequences are summarized in **Table 2**.

Western Blotting

Western blotting was performed as described in our previous experiments (Jiang et al., 2017; Zhao et al., 2017). The following primary antibodies were used in this study: GAPDH, p-JNK1/2, JNK1/2, p-STAT3, STAT3 (all from Cell Signaling Technology, Boston, United States), cyclin D1, BCL-xL (both from Abcam, Cambridge, United Kingdom), and ILF2 (from ZenBioScience, China).

EdU Incorporation Assay

The EdU (5-ethynyl-2′-deoxyuridine) incorporation assay was performed using a Cell-Light EdU DNA Cell Proliferation Kit (RiboBio, Guangzhou, China).

Flow Cytometry Analysis

Cell cycle and apoptosis analyses were performed by flow cytometry using a DNA content quantitation assay kit (Solarbio, Beijing, China) and an Annexin V-fluorescein isothiocyanate (FITC) apoptosis measurement kit (BD Biosciences, United States).

Enzyme-Linked Immunosorbent Assay

After 24 h of transfection as described above, cells were stimulated with M5, and culture supernatants were collected after 12 h. The secretion of IL-6 and IL-8 was measured using specific ELISA kits (Elabscience, China). The absorbance at 450 nm was measured with a microplate reader (BioTek, United States).

RNA Pull-Down Assays and Mass Spectrometry Analyses

The interaction between KLHDC7B-DT and RNA-binding protein was detected by a Pierce Magnetic RNA–Protein Pull-Down Kit (Thermo Fisher Scientific, United States). A Fast Digest Hind III Kit (Thermo Scientific) and MEGA Script Kit (Life Technologies) were used for digestion, linearization, and transcription. Biotin-labelled KLHDC7B-DT was incubated with total cell lysates of HaCaT cells, and eluted proteins were purified. The sequences of biotin-labelled KLHDC7B-DT are shown in **Supplementary Table S1**. Finally, interacting proteins were identified by mass spectrometry and Western blotting.

RNA Immunoprecipitation Assay

RIP was carried out using a Magna RIP Kit (Millipore, Billerica, MA, United States). HaCaT and Ker-CT cells were lysed with RNA lysis buffer, and cell lysates were incubated with RIP buffer containing magnetic beads conjugated to anti-ILF2 (ZenBioScience, #382994) or negative control IgG antibody (Millipore, Billerica, MA, United States) for 4 h at 4°C. Then, qRT–PCR and agarose electrophoresis were implemented to detect RNA. The sequences of primers for qRT–PCR are shown in **Supplementary Table S2**.

Immunohistochemistry

IHC was performed as previously described (Jiang et al., 2017; Zhao et al., 2017). Primary antibodies against ILF2

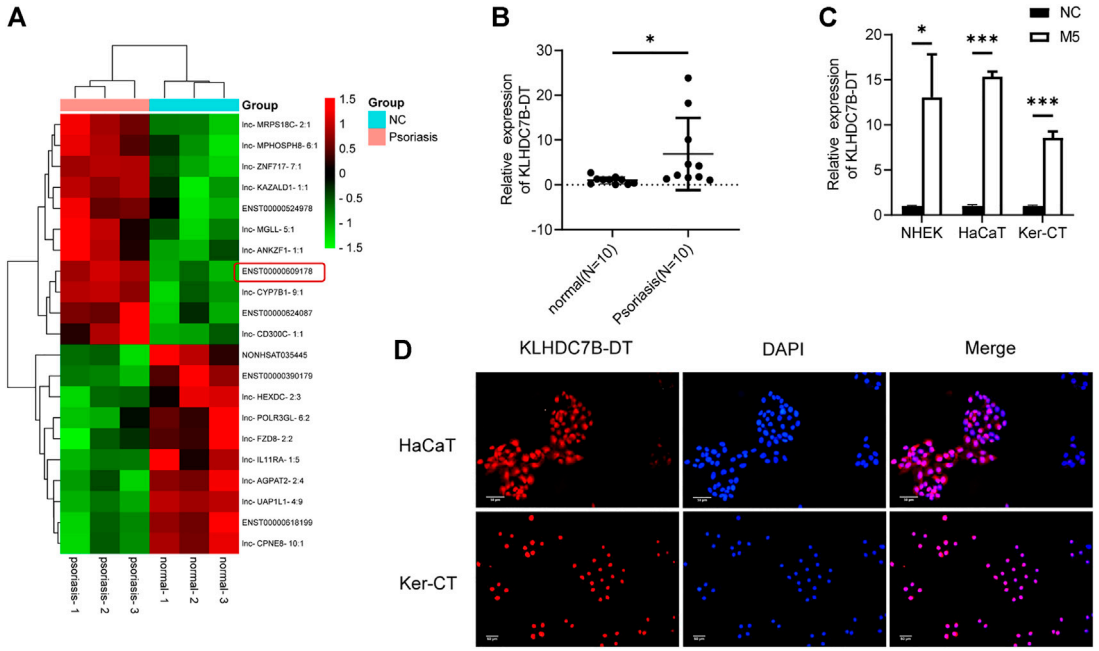


FIGURE 1 | KLHDC7B-DT is upregulated in psoriatic tissues and M5-induced keratinocytes. **(A)** Differentially expressed lncRNAs detected in psoriatic and normal skin tissues using a lncRNA microarray (n = 3, Accession: GSE181318). **(B)** The expression level of KLDHC7B-DT in psoriatic tissues and normal controls (n = 10). **(C)** The expression level of KLDHC7B-DT in M5-induced NHEKs, HaCaT cells, and Ker-CT cells. **(D)** Fluorescence *in situ* hybridization showed the expression and localization of KLDHC7B-DT in HaCaT and Ker-CT cells. A Cy3-labelled probe (red) was used to show KLDHC7B-DT. DAPI (blue) was performed to show the nuclei of cells. Data are shown as the means ± SD, $^{*}p < 0.05$, $^{***}p < 0.001$. All experiments were repeated at least three times. NHEK: normal human epidermal keratinocytes.

(ZenBioScience, #382994) were used in this study. IHC staining was quantified using H-scores, which incorporate the staining intensity and the percentage of positively-stained cells (range 0–100%).

Immunofluorescence

Immunofluorescence was performed according to the manufacturer's directions. Samples were analysed using a Zeiss LSM710 Confocal Laser Scanning Microscope (Carl Zeiss, Oberkochen, Germany).

Statistical Analysis

Data from at least three independent experiments are presented as the means ± standard deviation (SD). Statistical significance was analysed using GraphPad Prism (version 6, San Diego, United States). Differences between the two groups were analysed using Student's t test. $p < 0.05$ was considered statistically significant.

RESULTS

KLHDC7B-DT Is Upregulated in Psoriatic Tissues and M5-Induced Keratinocytes

To identify new lncRNAs associated with the pathogenesis of psoriasis, a lncRNA microarray was performed (n = 3, accession number: GSE181318). The microarray results showed that KLHDC7B-DT was one of the most significantly upregulated lncRNAs in psoriatic tissues, which was also confirmed by qRT–PCR (**Figures 1A,B**). Subsequently, an *in vitro* psoriasis model was established by stimulating keratinocytes with M5. As expected, the expression of KLHDC7B-DT was increased in M5-treated NHEKs, HaCaT cells, and Ker-CT cells (**Figure 1C**). FISH assays in HaCaT and Ker-CT cells indicated that KLHDC7B-DT was mainly distributed in the nucleus (**Figure 1D**). In addition, Sanger sequencing showed that the sequence of the amplified product was consistent with that of the template (**Supplementary Figure S2**). These results indicated that aberrantly upregulated KLHDC7B-DT expression might be associated with the pathogenesis of psoriasis.

KLHDC7B-DT Regulates Proliferation and Inflammatory Factor Secretion in M5-Induced Keratinocytes

We next investigated the potential biological roles of KLHDC7B-DT in M5-induced keratinocytes. qRT–PCR showed that KLHDC7B-DT expression was significantly suppressed by siRNA in HaCaT and Ker-CT cells (**Figure 2A**). EdU incorporation assays showed that M5 treatment increased EdU-positive cells, and the proliferation-promoting effect was inhibited by KLHDC7B-DT knockdown (**Figure 2B**). Cell cycle results indicated that M5 treatment led to a reduced proportion in

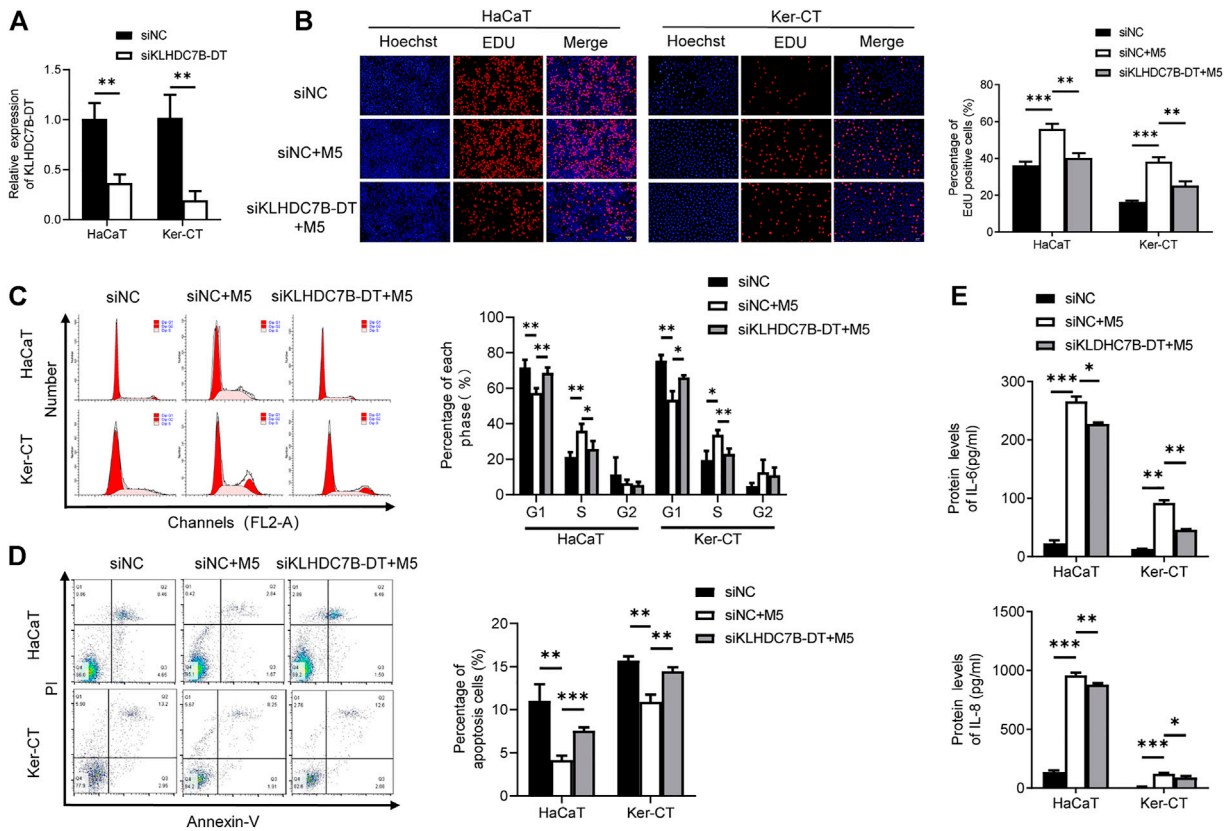


FIGURE 2 | KLHDC7B-DT regulates proliferation and inflammatory factor secretion in M5-induced keratinocytes. **(A)** The efficiency of KLDHC7B-DT knockdown was detected by qRT–PCR. **(B)** Proliferative effects were detected by EdU assay. **(C)** The cell cycle was detected by flow cytometry through PI staining. **(D)** Cell apoptosis was detected by flow cytometry through Annexin-V/PI staining. **(E)** Protein levels of IL-6 and IL-8 were detected by ELISA. Data are shown as the means ± SD, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. All experiments were repeated at least three times.

the G1 phase and an increased proportion in the S phase, and KLHDC7B-DT knockdown reversed the M5-induced effects (**Figure 2C**). Moreover, cell apoptosis analysis revealed that M5 treatment suppressed the proportion of early and late apoptotic cells, whereas KLHDC7B-DT knockdown induced a higher proportion of early and late apoptotic cells (**Figure 2D**). These findings suggested that KLHDC7B-DT regulates the proliferation of keratinocytes in psoriasis.

ELISAs showed that the M5-induced expression of IL-6 and IL-8 was partially reversed after KLHDC7B-DT knockdown in keratinocytes (**Figure 2E**). Collectively, our data demonstrated that KLHDC7B-DT knockdown reverses M5-induced hyperproliferation and the inflammatory response in keratinocytes, indicating that KLHDC7B-DT might play an important role in psoriasis by regulating the proliferation and inflammatory response of keratinocytes.

Overexpression of KLHDC7B-DT Facilitates Proliferation and Secretion of Inflammatory Factors in Keratinocytes

To further verify the role of KLHDC7B-DT in psoriasis, we stably overexpressed KLHDC7B-DT *via* transfection of a plasmid into M5-induced HaCaT and Ker-CT cells (**Figure 3A**). EdU assays showed that HaCaT and Ker-CT cells overexpressing KLHDC7B-DT had more EdU-positive cells (**Figure 3B**). The cell cycle results demonstrated that KLHDC7B-DT overexpression led to a reduced proportion of cells in the G1 phase and an increased proportion of cells in the S phase (**Figure 3C**). Moreover, cell apoptosis analysis revealed that KLHDC7B-DT overexpression inhibited the proportion of early and late apoptosis (**Figure 3D**). Overexpression of KLHDC7B-DT also significantly upregulated the IL-6 and IL-8 protein expression levels in M5-treated HaCaT and Ker-CT cells (**Figure 3E**). Taken together, these findings demonstrated that KLHDC7B-DT overexpression promotes the proliferation and inflammatory response in M5-induced keratinocytes.

KLHDC7B-DT Binds With the ILF2 Protein and Activates JNK/STAT3 Signalling

To further understand the molecular mechanism by which KLHDC7B-DT regulates proliferation and the inflammatory response in M5-induced keratinocytes, we precipitated synthesized KLHDC7B-DT *in vitro* and subjected it to mass spectrometric identification. RNA pull-down assays in

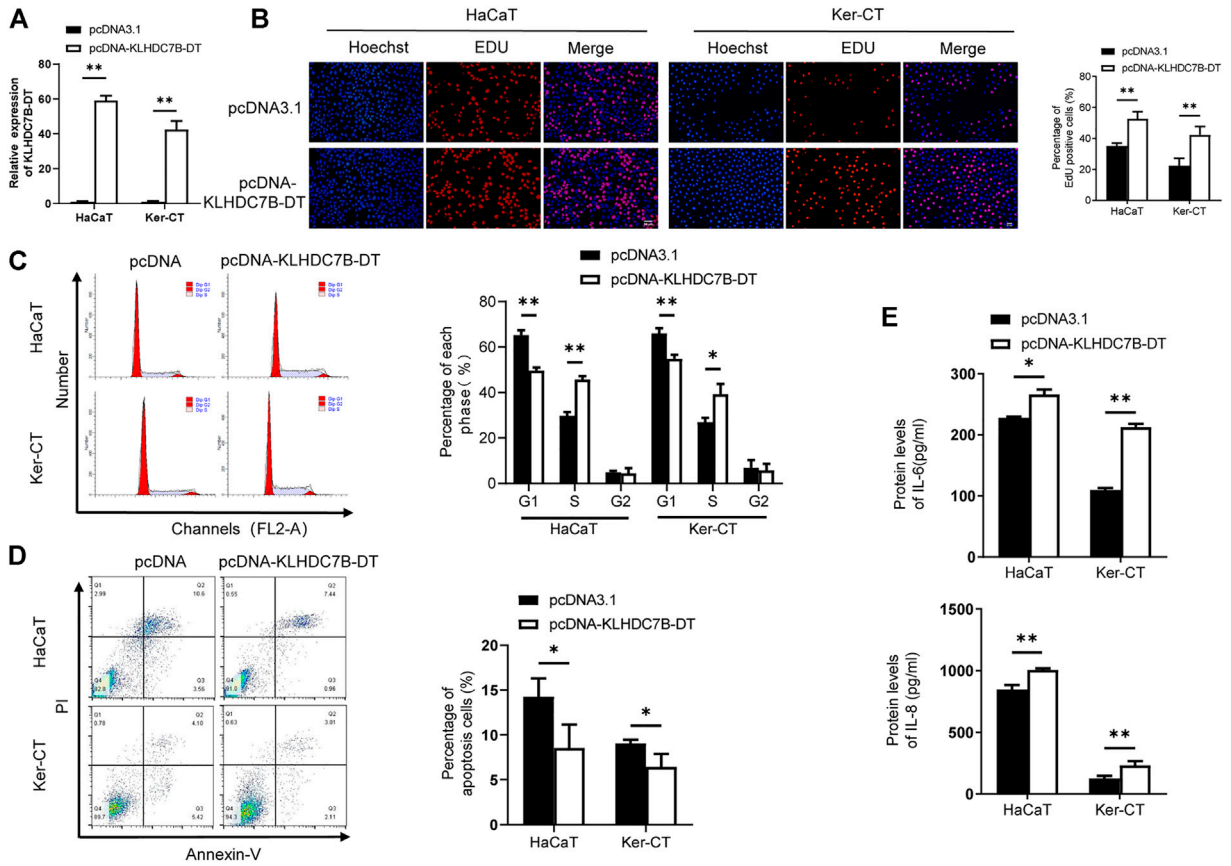


FIGURE 3 | The proliferative and inflammatory effects of KLDHC7B-DT overexpression in M5-induced keratinocytes. **(A)** The efficiency of KLDHC7B-DT overexpression detected by qRT–PCR. **(B)** Proliferative effects of KLDHC7B-DT overexpression on M5-treated keratinocytes detected by EdU assay. **(C)** The cell cycle was detected by flow cytometry through PI staining. **(D)** Cell apoptosis was detected by flow cytometry through Annexin-V/PI staining. **(E)** Protein levels of IL-6 and IL-8 were detected by ELISA. Data are shown as the means ± SD, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$. All experiments were repeated at least three times.

HaCaT cells showed that several proteins were identified as interacting partners of KLHDC7B-DT (**Figure 4A**). The binding proteins were further detected and analysed using mass spectrometry, which showed that KLHDC7B-DT could bind to the interacting proteins. Kyoto Encyclopedia of Genes and Genomes analysis suggested that these proteins were involved in the JAK-STAT and MAPK signalling pathways, while Gene Ontology analysis revealed the cell cycle, apoptosis, and inflammation functions of these proteins (**Supplementary Figure S3A,B**). Based on the mass spectrum assay (**Figure 4B**), we found that ILF2 was enriched in the KLHDC7B-DT-sense group among the detected proteins. Western blotting validated the existence of ILF2 within the RNA pull-down samples of KLHDC7B-DT in HaCaT cells (**Figure 4C**). In addition, RIP assays demonstrated the interaction between ILF2 and KLHDC7B-DT in HaCaT and Ker-CT cells (**Figure 4D**). By detecting the enriched RNA, the amplification band of KLHDC7B-DT was obtained by agarose electrophoresis (**Figure 4E**).

The expression of ILF2 was detected after KLHDC7B-DT knockdown and overexpression in HaCaT and Ker-CT cells, and no significant differences in ILF2 mRNA or protein expression were found (**Figures 4F,G**). To further determine whether ILF2 was involved in the upregulation of KLHDC7B-DT in psoriasis, ILF2 was knocked down in HaCaT and Ker-CT cells. The interference efficiency was verified by qRT–PCR (**Supplementary Figure S4**) and Western blotting (**Figure 4H**). Notably, ILF2 knockdown inhibited the expression of KLHDC7B-DT (**Figure 4I**).

Western blotting showed that the levels of p-STAT3 and p-JNK were significantly decreased after KLHDC7B-DT knockdown. In contrast, the protein levels of p-STAT3 and p-JNK were significantly increased after KLHDC7B-DT overexpression (**Figure 4J**). These results suggested that KLHDC7B-DT functioned by regulating the STAT3 and JNK signalling pathways in psoriasis.

ILF2 is Upregulated in Psoriatic Tissues and Contributes to the Proliferation and Secretion of Inflammatory Factors in Keratinocytes

Next, we investigated the role of ILF2 in psoriasis. qRT–PCR ($n = 3$) and Western blotting ($n = 3$) demonstrated that ILF2

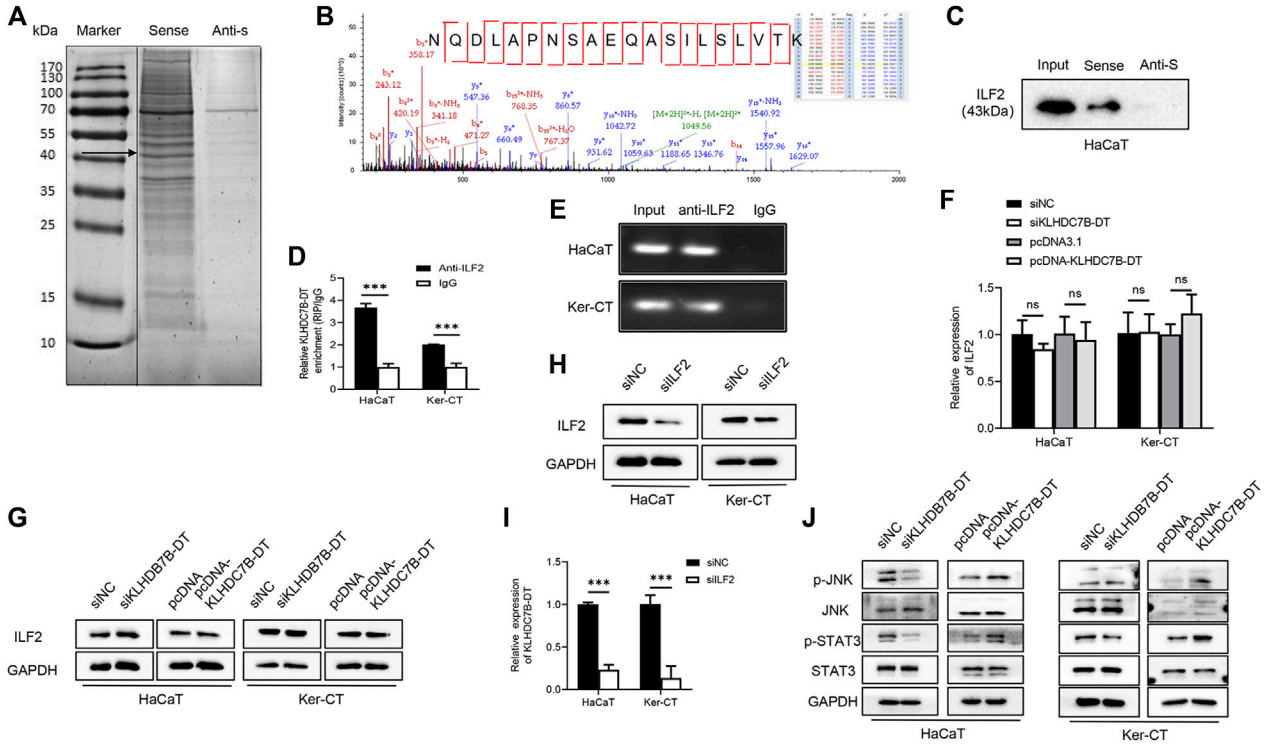


FIGURE 4 | KLHDC7B-DT can bind with ILF2 protein and activate JNK/STAT3 signalling. **(A)** Silver staining of interacting proteins after the KLDHC7B-DT pulldown assay. **(B)** Mass spectrometry of ILF2. **(C)** Western blotting was used to detect ILF2 in the RNA pulldown product. **(D)** RIP experiments were performed using an antibody against ILF2 on lysates from HaCaT and Ker-CT cells; IgG was used as the control group. **(E)** Agarose electrophoresis detected KLDHC7B-DT in products from RIP. **(F,G)** mRNA and protein levels of ILF2 in KLDHC7B-DT knockdown or overexpression keratinocytes. **(H)** The efficiency of ILF2 knockdown was detected by Western blotting. **(I)** The mRNA levels of KLDHC7B-DT after ILF2 knockdown. **(J)** Activation of the STAT/JNK pathway after KLDHC7B-DT knockdown or overexpression. Data are shown as the means ± SD, ***$p < 0.001$. All experiments were repeated at least three times.

was overexpressed in psoriatic tissues (**Figures 5A,B**). We carried out IHC staining to analyze the protein expression of ILF2 in 10-paired psoriatic tissues and normal tissues. The H-score indicated that ILF2 expression was increased in psoriatic tissues in contrast to normal tissues (**Figure 5C**). The expression of ILF2 was increased in M5-treated HaCaT and Ker-CT cells (**Figure 1D**). A cell immunofluorescence assay demonstrated that ILF2 was mainly localized in the nucleus of HaCaT cells and Ker-CT cells (**Figure 5D**). EdU assays showed that ILF2 knockdown inhibited the proliferation of keratinocytes (**Figure 6A**). Cell cycle analysis showed that ILF2 knockdown increased the proportion of the G1 phase and decreased the proportion of the S phase in keratinocytes (**Figure 6B**). Annexin-V/PI staining results demonstrated that the proportion of apoptotic cells increased after ILF2 knockdown (**Figure 6C**). In addition, ELISA analysis showed that the increased protein levels of IL-6 and IL-8 in M5-induced keratinocytes were significantly decreased after ILF2 knockdown (**Figure 6D**). Collectively, these findings showed that ILF2 knockdown inhibits the proliferation, cell cycle and secretion of inflammatory factors and promotes apoptosis in M5-induced keratinocytes.

ILF2 Promotes Hyperproliferation and Inflammation Through STAT3/JNK Pathways in a KLHDC7B-DT-Dependent Manner

To explore the regulatory mechanism of ILF2 and KLHDC7B-DT in psoriasis, ILF2 siRNA was transfected alone or together with pcDNA-KLHDC7B-DT into M5-treated HaCaT and Ker-CT cells. Then, the proliferation, cell cycle, apoptosis, and inflammatory factor secretion of keratinocytes were detected. The reduced percentage of EdU-positive cells and G1/S growth arrest caused by ILF2 knockdown could be rescued by KLHDC7B-DT overexpression (**Figures 6A,B**). ILF2 knockdown increased the proportion of apoptotic cells, which was significantly reversed by overexpression of KLHDC7B-DT (**Figure 6C**). Moreover, the ILF2 knockdown-mediated decrease in IL-6 and IL-8 protein levels was significantly reversed by KLHDC7B-DT overexpression (**Figure 6D**).

Then, we stably overexpressed ILF2 *via* transfection of a plasmid into M5-induced HaCaT cells. The overexpression efficiency was verified by Western blotting (**Supplementary Figure S5**). Then pcDNA-ILF2 was transfected alone or together with siKLHDC7B-DT into M5-treated HaCaT cells, the proliferation, cell cycle, apoptosis, and inflammatory factor

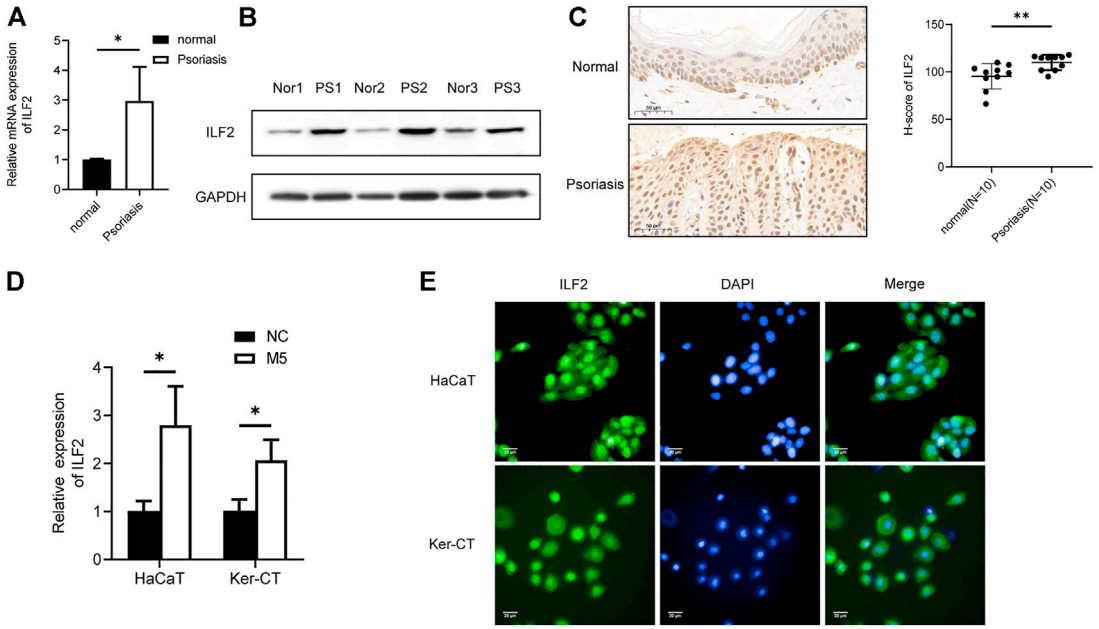


FIGURE 5 | ILF2 is highly expressed in psoriatic lesions. **(A)** The mRNA expression level of ILF2 in psoriatic tissues ($n = 3$). **(B)** The protein expression level of ILF2 in psoriatic tissues ($n = 3$). **(C)** The expression of ILF2 in normal and psoriatic tissues was detected by IHC ($n = 10$). The H-score was calculated based on the percentage of ILF2+ cells and staining intensity.**(D)** The expression level of ILF2 in M5-treated HaCaT and Ker-CT cells. **(E)** Subcellular localization of ILF2 in HaCaT and Ker-CT cells. Data are shown as the means ± SD, $^*p < 0.05$. All experiments were repeated at least three times.

secretion of keratinocytes were detected. The increased percentage of EdU-positive cells and promotion of G1/S transition caused by ILF2 overexpression could be reversed by KLHDC7B-DT knockdown (**Supplementary Figure S6A,B**). ILF2 overexpression decreased the proportion of apoptotic cells, which was significantly rescued by KLHDC7B-DT knockdown (**Supplementary Figure S6C**). Moreover, the ILF2 overexpression-mediated increasein IL-6 and IL-8 protein levels was significantly reversed by KLHDC7B-DT knockdown (**Supplementary Figure S6D**).

As it has been reported that ILF2 is involved in inflammatory responses through the STAT3 and JNK signalling pathways, we investigated whether ILF2 induces phosphorylation of STAT3/JNK in HaCaT and Ker-CT cells. Interestingly, ILF2 knockdown decreased the phosphorylation levels of STAT3/JNK. More importantly, decreased STAT3/JNK phosphorylation after ILF2 knockdown could be reversed by KLHDC7B-DT overexpression (**Figure 6E**). Taken together, the results indicated that ILF2 promotes hyperproliferation and inflammation through STAT3/JNK pathways in a KLHDC7B-DT-dependent manner.

DISCUSSION

Psoriasis is a chronic inflammatory skin disorder (Nickoloff et al., 2007b; Bos, 2007; Boehncke and Sterry, 2009). Keratinocytes actively participate in the initiation and maintenance of psoriatic skin inflammation through the production of various proinflammatory cytokines (Lowes et al., 2014). In this study, we performed a lncRNA microarray, and the results showed that KLHDC7B-DT was one of the most significantly upregulated lncRNAs in psoriatic tissues. Then, we verified the results in psoriatic lesions and a psoriasis model *in vitro* by stimulating keratinocytes with IL-17A, IL-22, IL-1α, oncostatin M, and TNF-α (M5) (Rabeony et al., 2014; Li et al., 2019). We found that the expression levels of KLHDC7B-DT, as well as its binding protein ILF2, were upregulated in both skin lesions and psoriasis model keratinocytes, suggesting that KLHDC7B-DT and ILF2 might be involved in the pathogenesis of psoriasis.

LncRNAs have been revealed to participate in the pathogenesis of psoriasis, including MEG3, GAS5, and MIR31HG (Gao et al., 2018; Jia et al., 2019; Ahmed Shehata et al., 2021). Here, we demonstrated that KLHDC7B-DT positively regulates the proliferation of keratinocytes. Cell cycle analysis and Annexin-V/PI staining indicated that KLHDC7B-DT regulates the G1/S transition and cell apoptosis. These results suggested that KLHDC7B-DT might be involved in the pathogenesis of psoriasis by regulating keratinocyte proliferation and apoptosis.

IL-6 and IL-8 have been proven to play important roles in psoriasis (Elder et al., 1992; Yoshinaga et al., 1995). Previous studies also demonstrated that serum levels of proinflammatory cytokines reflect disease activity and treatment response (Grossman et al., 1989; Biasi et al., 1998). In this study, we found that the expression of IL-6 and IL-8 was regulated by KLHDC7B-DT in keratinocytes. Our findings suggested that KLHDC7B-DT contributes to psoriatic skin inflammation by regulating the expression of IL-6 and IL-8.

It has become increasingly clear that the biological functions of lncRNAs are associated with their unique subcellular localization (Carlevaro-Fita and Johnson, 2019). LncRNAs can function by

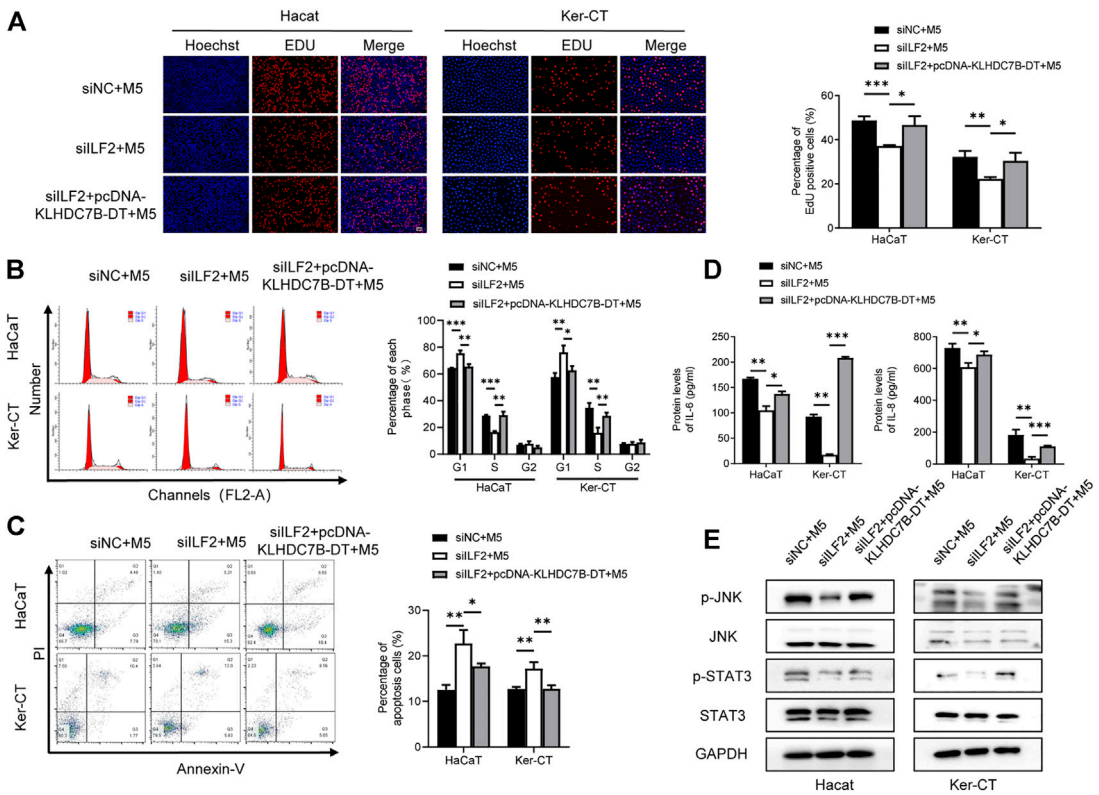


FIGURE 6 | ILF2 promotes hyperproliferation and inflammation through STAT3/JNK pathways in M5-induced keratinocytes. **(A)** Using an EdU assay, the proliferative effects of M5-induced keratinocytes were detected after transfection with siILF2 and siILF2+pcDNA-KLDHC7B-DT. **(B)** The cell cycle was detected by flow cytometry through PI staining. **(C)** Cell apoptosis was detected by flow cytometry through Annexin-V/PI staining. **(D)** Protein levels of IL-6 and IL-8 were detected by ELISA. **(E)** Activation of the STAT/JNK pathway detected by Western blotting. Data are shown as the means ± SD, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. All experiments were repeated at least three times.

interacting with proteins *via* structural interactions and/or complementary base pairing (Tang et al., 2020). In this study, both ILF2 and KLHDC7B-DT were found to be mainly distributed in the nucleus. Furthermore, we confirmed the interaction between KLHDC7B-DT and ILF2 protein by RNA pull-down and RIP assays. We found that ILF2 acted as the upstream regulator of KLHDC7B-DT and positively regulated its expression. In addition, previous studies reported that ILF2 was involved in inflammatory responses through the STAT3 and JNK inflammatory signalling pathways (Jia et al., 2020). We confirmed that ILF2 was upregulated in psoriatic tissues and M5-induced keratinocytes. Moreover, overexpression of KLHDC7B-DT rescued the effects of ILF2 knockdown on keratinocyte proliferation and cytokine secretion. In contrast, KLHDC7B-DT konckdown reversed the effects of ILF2 overexpression. These results suggested that ILF2 regulates proliferation and the inflammatory response *via* KLHDC7B-DT in keratinocytes.

Furthermore, KEGG and GO analyses showed that proteins binding with KLDHC7B-DT were involved in the JAK-STAT and MAPK signalling pathways. Previous studies have reported that STAT3 and JNK play crucial roles in regulating immune responses and participate in the pathogenesis of psoriasis (Sano et al., 2005; Liang et al., 2019). IL-6 and IL-8 have been proven to play important roles in psoriasis, and both of them can initiate STAT3 and JNK MAPK pathway, thus involving in regulating cell inflammation and proliferation of keratinocytes (Mori et al., 2011; Chen et al., 2015; Fu et al., 2015). Furthermore, KLHDC7B-DT was found to directly bind to the IL-6 promoter activate IL-6 transcription, up-regulate IL-6 expression. Through enhancing IL-6 secretion, KLHDC7B-DT activated STAT3 signaling in PDAC cells (Jia et al., 2020). In addition to the STAT3 signaling pathway, IL-6 also activates the MAPK cascade (Neurath and Finotto, 2011). Our findings showed that KLHC7B-DT is closely related to the activation of the STAT3 and JNK signalling pathways, and indicated that KLHDC7B-DT activated STAT3 and JNK signaling pathways by promoting IL-6 and IL-8 secretion. Furthermore, overexpression of KLHDC7B-DT rescued the expression of p-STAT3 and p-JNK after ILF2 knockdown, which indicated that ILF2 regulates the STAT3/JNK pathways *via* KLHDC7B-DT. To sum up, our results suggested that ILF2 regulates proliferation and inflammatory reactions *via* the KLHDC7B-DT and STAT3/JNK signalling pathways in keratinocytes.

Although various studies have emphasized the important role of lncRNAs, few have focused on the underlying mechanism driving the abnormal expression of lncRNAs in psoriasis. In the present study, our data demonstrated that upregulated expression of KLHDC7B-DT was associated with high levels of ILF2 in psoriatic tissues.

However, this study was subject to some limitations. Our sample sizes were relatively small, and further investigations with larger samples are required to verify our findings. The expression of KLHDC7B-DT and ILF2 was focused only on psoriasis. Other inflammatory diseases, including atopic dermatitis, should be included to investigate the specificity of the biomarkers.

CONCLUSION

In conclusion, we demonstrated that KLHDC7B-DT and ILF2 are overexpressed in psoriatic skin lesions. Overexpression of KLHDC7B-DT in psoriasis was related to the aberrant expression of ILF2. ILF2 and KLHDC7B-DT were involved in the hyperproliferation of keratinocytes and skin inflammation in psoriasis. In addition, we verified that ILF2 functions in a KLHDC7B-DT-dependent manner.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The name of the repository and accession number can be found below: GEO, NCBI; GSE181318.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Shandong University, Qilu Hospital. Written informed consent to participate in this study was provided by the participants' legal guardian/ next of kin.

AUTHOR CONTRIBUTIONS

XY contributed to the conception, design of the work, analysis, and interpretation of data and drafted this manuscript. ZY, MZ, and CC contributed to the design of the work, analysis, and interpretation of the data. SH, XL, HZ, and HW contributed to the acquisition and interpretation of the data. QS, XY, and JY contributed to the conception and design of the work. All authors read and approved the final manuscript.

FUNDING

This work was supported by National Natural Science Foundation of China (Grant Nos. 81972937, 82173419, and 82003344), Natural Science Foundation of Shandong Province (ZR2020QH137 and ZR2021MH367).

ACKNOWLEDGMENTS

We would like to acknowledge the helpful comments on this article received from our reviewers. We thank the staff of the laboratory of Basic Medical Sciences for their help during the experiment.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.890624/full#supplementary-material

REFERENCES

Ahmed Shehata, W., Maraee, A., Abd El Monem Ellaithy, M., Tayel, N., Abo-Ghazala, A., and Mohammed El-Hefnawy, S. (2021). Circulating Long Noncoding RNA Growth Arrest-specific Transcript 5 as a Diagnostic Marker and Indicator of Degree of Severity in Plaque Psoriasis. *Int. J. Dermatol.* 60 (8), 973–979. doi:10.1111/ijd.15494

Albanesi, C., De Pità, O., and Girolomoni, G. (2007). Resident Skin Cells in Psoriasis: a Special Look at the Pathogenetic Functions of Keratinocytes. *Clin. Dermatol.* 25 (6), 581–588. doi:10.1016/j.clindermatol.2007.08.013

Albanesi, C., Scarponi, C., Giustizieri, M., and Girolomoni, G. (2005). Keratinocytes in Inflammatory Skin Diseases. *Cdtia* 4 (3), 329–334. doi:10.2174/1568010054022033

Biasi, D., Carletto, A., Caramaschi, P., Bellavite, P., Maleknia, T., Scambi, C., et al. (1998). Neutrophil Functions and IL-8 in Psoriatic Arthritis and in Cutaneous Psoriasis. *Inflammation* 22 (5), 533–543. doi:10.1023/a:1022354212121

Boehncke, W.-H., and Schön, M. P. (2015). Psoriasis. *The Lancet* 386 (9997), 983–994. doi:10.1016/s0140-6736(14)61909-7

Boehncke, W.-H., and Sterry, W. (2009). Psoriasis--a Systemic Inflammatory Disorder: Clinic, Pathogenesis and Therapeutic Perspectives. *J. Dtsch Dermatol. Ges.* 7 (11), 946–952. doi:10.1111/j.1610-0387.2009.07083.x

Bos, J. D. (2007). Psoriasis, Innate Immunity, and Gene Pools. *J. Am. Acad. Dermatol.* 56 (3), 468–471. doi:10.1016/j.jaad.2006.12.006

Carlevaro-Fita, J., and Johnson, R. (2019). Global Positioning System: Understanding Long Noncoding RNAs through Subcellular Localization. *Mol. Cel.* 73 (5), 869–883. doi:10.1016/j.molcel.2019.02.008

Chen, W.-C., Tseng, C.-K., Chen, Y.-H., Lin, C.-K., Hsu, S.-h., Wang, S.-N., et al. (2015). HCV NS5A Up-Regulates COX-2 Expression via IL-8-Mediated Activation of the ERK/JNK MAPK Pathway. *PLoS ONE* 10 (7), e0133264. doi:10.1371/journal.pone.0133264

Du, H., Le, Y., Sun, F., Li, K., and Xu, Y. (2019). ILF2 Directly Binds and Stabilizes CREB to Stimulate Malignant Phenotypes of Liver Cancer Cells. *Anal. Cel Pathol (Amst)* 2019, 1575031. doi:10.1155/2019/1575031

Elder, J. T., Sartor, C. I., Boman, D. K., Benrazavi, S., Fisher, G. J., and Pittelkow, M. R. (1992). Interleukin-6 in Psoriasis: Expression and Mitogenicity Studies. *Arch. Dermatol. Res.* 284 (6), 324–332. doi:10.1007/bf00372034

Fatica, A., and Bozzoni, I. (2014). Long Non-coding RNAs: New Players in Cell Differentiation and Development. *Nat. Rev. Genet.* 15 (1), 7–21. doi:10.1038/nrg3606

Fu, X.-T., Dai, Z., Song, K., Zhang, Z.-J., Zhou, Z.-J., Zhou, S.-L., et al. (2015). Macrophage-secreted IL-8 Induces Epithelial-Mesenchymal Transition in Hepatocellular Carcinoma Cells by Activating the JAK2/STAT3/Snail Pathway. *Int. J. Oncol.* 46 (2), 587–596. doi:10.3892/ijo.2014.2761

Gao, J., Chen, F., Hua, M., Guo, J., Nong, Y., Tang, Q., et al. (2018). Knockdown of lncRNA MIR31HG Inhibits Cell Proliferation in Human HaCaT Keratinocytes. *Biol. Res.* 51 (1), 30. doi:10.1186/s40659-018-0181-8

Grossman, R. M., Krueger, J., Yourish, D., Granelli-Piperno, A., Murphy, D. P., May, L. T., et al. (1989). Interleukin 6 Is Expressed in High Levels in Psoriatic Skin and Stimulates Proliferation of Cultured Human Keratinocytes. *Proc. Natl. Acad. Sci. U.S.A.* 86 (16), 6367–6371. doi:10.1073/pnas.86.16.6367

Huang, W., Liu, J., Yan, J., Huang, Z., Zhang, X., Mao, Y., et al. (2020). LncRNA LINC00470 Promotes Proliferation through Association with NF45/NF90 Complex in Hepatocellular Carcinoma. *Hum. Cel.* 33 (1), 131–139. doi:10.1007/s13577-019-00288-8

Iyer, M. K., Niknafs, Y. S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., et al. (2015). The Landscape of Long Noncoding RNAs in the Human Transcriptome. *Nat. Genet.* 47 (3), 199–208. doi:10.1038/ng.3192

Jia, H.-Y., Zhang, K., Lu, W.-J., Xu, G.-W., Zhang, J.-F., and Tang, Z.-L. (2019). LncRNA MEG3 Influences the Proliferation and Apoptosis of Psoriasis Epidermal Cells by Targeting miR-21/caspase-8. *BMC Mol. Cel Biol* 20 (1), 46. doi:10.1186/s12860-019-0229-9

Jia, Y., Li, X., Nan, A., Zhang, N., Chen, L., Zhou, H., et al. (2020). Circular RNA 406961 Interacts with ILF2 to Regulate PM2.5-induced Inflammatory Responses in Human Bronchial Epithelial Cells via Activation of STAT3/JNK Pathways. *Environ. Int.* 141, 105755. doi:10.1016/j.envint.2020.105755

Jiang, M., Ma, W., Gao, Y., Jia, K., Zhang, Y., Liu, H., et al. (2017). IL-22-induced miR-122-5p Promotes Keratinocyte Proliferation by Targeting Sprouty2. *Exp. Dermatol.* 26 (4), 368–374. doi:10.1111/exd.13270

Kuwano, Y., Kim, H. H., Abdelmohsen, K., Pullmann, R., Martindale, J. L., Yang, X., et al. (2008). MKP-1 mRNA Stabilization and Translational Control by RNA-Binding Proteins HuR and NF90. *Mol. Cel Biol* 28 (14), 4562–4575. doi:10.1128/mcb.00165-08

Li, C., Xiao, L., Jia, J., Li, F., Wang, X., Duan, Q., et al. (2019). Cornulin Is Induced in Psoriasis Lesions and Promotes Keratinocyte Proliferation via Phosphoinositide 3-Kinase/Akt Pathways. *J. Invest. Dermatol.* 139 (1), 71–80. doi:10.1016/j.jid.2018.06.184

Li, M.-x., Wang, H.-y., Yuan, C.-h., Ma, Z.-l., Jiang, B., Li, L., et al. (2021). KLHDC7B-DT Aggravates Pancreatic Ductal Adenocarcinoma Development via Inducing Cross-Talk between Cancer Cells and Macrophages. *Clin. Sci.* 135 (4), 629–649. doi:10.1042/cs20201259

Liang, J., Chen, P., Li, C., Li, D., Wang, J., Xue, R., et al. (2019). IL-22 Down-Regulates Cx43 Expression and Decreases Gap Junctional Intercellular Communication by Activating the JNK Pathway in Psoriasis. *J. Invest. Dermatol.* 139 (2), 400–411. doi:10.1016/j.jid.2018.07.032

Lin, A. M., Rubin, C. J., Khandpur, R., Wang, J. Y., Riblett, M., Yalavarthi, S., et al. (2011). Mast Cells and Neutrophils Release IL-17 through Extracellular Trap Formation in Psoriasis. *J. Immunol.* 187 (1), 490–500. doi:10.4049/jimmunol.1100123

Lowes, M. A., Suárez-Fariñas, M., and Krueger, J. G. (2014). Immunology of Psoriasis. *Annu. Rev. Immunol.* 32, 227–255. doi:10.1146/annurev-immunol-032713-120225

Mirza, A. H., Berthelsen, C. H., Seemann, S. E., Pan, X., Frederiksen, K. S., Vilien, M., et al. (2015). Transcriptomic Landscape of lncRNAs in Inflammatory Bowel Disease. *Genome Med.* 7 (1), 39. doi:10.1186/s13073-015-0162-2

Mori, T., Miyamoto, T., Yoshida, H., Asakawa, M., Kawasumi, M., Kobayashi, T., et al. (2011). IL-1 and TNF -initiated IL-6-STAT3 Pathway Is Critical in Mediating Inflammatory Cytokines and RANKL Expression in Inflammatory Arthritis. *Int. Immunol.* 23 (11), 701–712. doi:10.1093/intimm/dxr077

Neurath, M. F., and Finotto, S. (2011). IL-6 Signaling in Autoimmunity, Chronic Inflammation and Inflammation-Associated Cancer. *Cytokine Growth Factor. Rev.* 22 (2), 83–89. doi:10.1016/j.cytogfr.2011.02.003

Nickoloff, B. J., Qin, J.-Z., and Nestle, F. O. (2007a). Immunopathogenesis of Psoriasis. *Clinic Rev. Allerg Immunol.* 33, 45–56. doi:10.1007/s12016-007-0039-2

Nickoloff, B. J., Xin, H., Nestle, F. O., and Qin, J.-Z. (2007b). The Cytokine and Chemokine Network in Psoriasis. *Clin. Dermatol.* 25 (6), 568–573. doi:10.1016/j.clindermatol.2007.08.011

Qiao, M., Li, R., Zhao, X., Yan, J., and Sun, Q. (2018). Up-regulated lncRNA-Msx2p1 Promotes the Growth of IL-22-stimulated Keratinocytes by Inhibiting miR-6731-5p and Activating S100A7. *Exp. Cel. Res.* 363 (2), 243–254. doi:10.1016/j.yexcr.2018.01.014

Rabeony, H., Petit-Paris, I., Garnier, J., Barrault, C., Pedretti, N., Guilloteau, K., et al. (2014). Inhibition of Keratinocyte Differentiation by the Synergistic Effect of IL-17A, IL-22, IL-1α, TNFα and Oncostatin M. *PloS one* 9 (7), e101937. doi:10.1371/journal.pone.0101937

Sano, S., Chan, K. S., Carbajal, S., Clifford, J., Peavey, M., Kiguchi, K., et al. (2005). Stat3 Links Activated Keratinocytes and Immunocytes Required for Development of Psoriasis in a Novel Transgenic Mouse Model. *Nat. Med.* 11 (1), 43–49. doi:10.1038/nm1162

Santoro, M., Nociti, V., Lucchini, M., De Fino, C., Losavio, F. A., and Mirabella, M. (2016). Expression Profile of Long Non-coding RNAs in Serum of Patients with Multiple Sclerosis. *J. Mol. Neurosci.* 59 (1), 18–23. doi:10.1007/s12031-016-0741-8

Srivastava, A., Nikamo, P., Lohcharoenkal, W., Li, D., Meisgen, F., Xu Landén, N., et al. (2017). MicroRNA-146a Suppresses IL-17-mediated Skin Inflammation and Is Genetically Associated with Psoriasis. *J. Allergy Clin. Immunol.* 139 (2), 550–561. doi:10.1016/j.jaci.2016.07.025

Szegedi, K., Sonkoly, E., Nagy, N., Németh, I. B., Bata-Csörgő, Z., Kemény, L., et al. (2010). The Anti-apoptotic Protein G1P3 Is Overexpressed in Psoriasis and Regulated by the Non-coding RNA, PRINS. *Exp. Dermatol.* 19 (3), 269–278. doi:10.1111/j.1600-0625.2010.01066.x

Tang, L., Liang, Y., Xie, H., Yang, X., and Zheng, G. (2020). Long Non-coding RNAs in Cutaneous Biology and Proliferative Skin Diseases: Advances and Perspectives. *Cell Prolif* 53 (1), e12698. doi:10.1111/cpr.12698

Wu, G.-C., Pan, H.-F., Leng, R.-X., Wang, D.-G., Li, X.-P., Li, X.-M., et al. (2015). Emerging Role of Long Noncoding RNAs in Autoimmune Diseases. *Autoimmun. Rev.* 14 (9), 798–805. doi:10.1016/j.autrev.2015.05.004

Yan, J. J., Qiao, M., Li, R. H., Zhao, X. T., Wang, X. Y., and Sun, Q. (2019). Downregulation of miR-145-5p Contributes to Hyperproliferation of Keratinocytes and Skin Inflammation in Psoriasis. *Br. J. Dermatol.* 180 (2), 365–372. doi:10.1111/bjd.17256

Yoshinaga, Y., Higaki, M., Terajima, S., Ohkubo, E., Nogita, T., Miyasaka, N., et al. (1995). Detection of Inflammatory Cytokines in Psoriatic Skin. *Arch. Dermatol. Res.* 287 (2), 158–164. doi:10.1007/bf01262325

Zhang, X., Bustos, M. A., Gross, R., Ramos, R. I., Takeshima, T. L., Mills, G. B., et al. (2021). Interleukin Enhancer-Binding Factor 2 Promotes Cell Proliferation and DNA Damage Response in Metastatic Melanoma. *Clin. Transl Med.* 11 (10), e608. doi:10.1002/ctm2.608

Zhao, G., Shi, L., Qiu, D., Hu, H., and Kao, P. N. (2005). NF45/ILF2 Tissue Expression, Promoter Analysis, and Interleukin-2 Transactivating Function. *Exp. Cel. Res.* 305 (2), 312–323. doi:10.1016/j.yexcr.2004.12.030

Zhao, X., Li, R., Qiao, M., Yan, J., and Sun, Q. (2017). MiR-548a-3p Promotes Keratinocyte Proliferation Targeting PPP3R1 after Being Induced by IL-22. *Inflammation* 41 (2), 496–504. doi:10.1007/s10753-017-0705-3

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.

The handling editor YS declared a shared affiliation with the author MZ at the time of review.

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.

Copyright © 2022 Yin, Yang, Zhu, Chen, Huang, Li, Zhong, Wen, Sun, Yu and Yan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

CASE REPORT
published: 09 June 2022
doi: 10.3389/fgene.2022.873635

A Novel Variant of X-Linked Moesin Gene in a Boy With Inflammatory Bowel Disease Like Disease-A Case Report

Youhong Fang[1], *Youyou Luo*[1], *Yang Liu*[2] *and Jie Chen*[1]*

[1]Department of Gastroenterology, The Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, China, [2]Department of Pathology, The Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, China

Variants in the *MSN* gene were recently reported as the cause of a primary immunodeficiency disease called X-linked moesin-associated immunodeficiency (X-MAID). Hitherto, only 10 patients were reported worldwide. Here, we report a boy who presented with recurrent high fever, oral ulcers, abdominal pain, and hematochezia for over 2 weeks. His serum inflammatory markers were elevated, and colonoscopy showed multiple colon ulcers and terminal ileum ulcers which resemble colitis caused by inflammatory bowel disease. A novel heterozygous variant c.934G>T(p.Glu312Ter) in the *MSN* gene was identified using whole exome sequencing (WES) and trio analysis. Intestinal ulcers were almost healed after inducing therapy with steroids and maintenance treatment of anti-TNFα therapy. We summarized the genotype and phenotype of reported X-MAID patients and presented the patient's unique phenotype in this study. This study also expanded the spectrum of *MSN* mutation-caused immunodeficiency.

Keywords: moesin gene, intestinal ulcer, colitis, very early-onset inflammatory bowel disease, primary immunodeficiency disease

OPEN ACCESS

Edited by:
Babak Behnam,
National Sanitation Foundation International, United States

Reviewed by:
Anthony Vandersteen,
Izaak Walton Killam Health Centre, Canada
David Bunyan,
Salisbury NHS Foundation Trust, United Kingdom
Dror S. Shouval,
Schneider Children's Medical Center, Israel

***Correspondence:**
Jie Chen
6185020@zju.edu.cn

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: *11 February 2022*
Accepted: *27 April 2022*
Published: *09 June 2022*

Citation:
Fang Y, Luo Y, Liu Y and Chen J (2022) A Novel Variant of X-Linked Moesin Gene in a Boy With Inflammatory Bowel Disease Like Disease-A Case Report. Front. Genet. 13:873635. doi: 10.3389/fgene.2022.873635

INTRODUCTION

Variants in the moesin (*MSN*) gene were first recognized as a cause of congenital disease in humans in 2016; the resulting disease is referred to as X-linked moesin-associated immunodeficiency (X-MAID, #300988) (Lagresle-Peyrou et al., 2016). Patients with hemizygous mutations in the *MSN* gene present with lymphopenia, hypogammaglobulinemia, and susceptibility to bacterial and viral infections (Lagresle-Peyrou et al., 2016). X-MAID patients also exhibit impaired T-cell proliferation and altered migration and adhesion capacities (Lagresle-Peyrou et al., 2016). MSN, ezrin and radixin are members of the ezrin-radixin-moesin (ERM) family. ERM is a cell-structure-related protein that regulates the actin cytoskeleton and plasma membranes (Burns et al., 2017). Although MSN, ezrin, and radixin are structurally and functionally similar, they demonstrate different physical functions and expression patterns. Ezrin protein is mostly expressed in epithelial cells, whereas MSN predominates in endothelial cells (Berryman et al., 1993). MSN and ezrin play important roles in lymphoid cell proliferation, migration and adhesion (Lagresle-Peyrou et al., 2016). *MSN* knockout mice show impaired chemotaxis of T and B lymphocytes (Hirata et al., 2012).

X-MAID is classified as a primary immunodeficiency (PID), and most PIDs are monogenic diseases. In recent years, PIDs have been reported to have a close relationship with very early-onset inflammatory bowel disease (VEOIBD), which is defined as inflammatory bowel disease (IBD) that

occurs before 6 years of age. PIDs that mainly present with IBD-like features belong to monogenic IBD as well. To date, more than 70 genes have been discovered that are associated with monogenic IBD. Some genes reported as associated with immune deficiency or inborn errors of immunity. The majority (>70–80%) of VEOBD patients not have a specific identified causal genetic etiology (Charbit-Henrion et al., 2018). However, the rate of genetic discoveries among VEOIBD patients is increasing, especially in the group of infantile onset IBD patients. The well-known genes related to monogenic IBD are *IL10*/*IL10R*, *XIAP* and *CGD*. Among these genes, the most common variants that lead to VEOIBD are *IL10R* mutations. The main mechanisms related to monogenic IBD include intestinal epithelial barrier function, phagocytic bactericidal activity, hyper- or autoimmune inflammatory pathways, and the development and function of the adaptive immune system (Kelsen et al., 2020). The relationship between PIDs and IBD-like disease is not completely known. Variants of the *MSN* gene have not been reported to be associated with intestinal colitis. Here, we report a VEOIBD patient with a novel variant of the X-linked *MSN* gene.

CASE DESCRIPTION

A boy aged 5 years and 2 months presented with recurrent high fever for 18 days, accompanied by a mild cough, abdominal pain, diarrhea and hematochezia. Although he was actively treated with intravenous antibiotics and immunoglobin, his fever was persistent. He was the first child in his family. A single episode of mild pneumonia had occurred when he was 4 months old and he received intravenous antibiotics for several days. There was no history of recurrent respiratory infection or eczema. No history of prolonged umbilical stump as well. He did not have chickenpox and was not vaccinated against this. On physical examination, he had moderate malnutrition, multiple ulcers in the oral cavity and on the surface of the tongue, and no lymphadenopathy or hepatosplenomegaly. Laboratory examinations revealed obvious elevation of inflammatory markers; the white blood cell count was 19.44×10^9/L, of which neutrophils accounted for 66.4%; the C-reactive protein (CRP) level was 17.34 mg/L; the erythrocyte sedimentation rate (ESR) was 73 mm/h; and the serum procalcitonin level was 0.412 ng/ml. In addition, the patient had moderate anemia, with a hemoglobin level of 84 g/L, and his serum albumin level was decreased, at 27.4 g/L. Serum antibodies against cytomegalovirus (CMV) and Epstein–Barr virus (EBV) were negative, and a T-SPOT TB test was negative. Number of $CD19^+$ B lymphocytes was reduced, accounting for 16.45% (reference range (r.r): 18.5–28.0%) of lymphocytes; $CD3^+$ T lymphocytes accounted for 54.1% (r.r: 56.0–68.0%) of lymphocytes, $CD4^+$ T lymphocytes accounted for 31.45% (r.r: 29.0–40.0%), and $CD8^+$ T lymphocytes accounted for 18.35% (r.r: 19.0–25.0%). The ratio of $CD4^+$ to $CD8^+$ T cells was normal, 1.71 (r.r: 1.1–2.0/1). Abdominal computed tomography scan showed irregular wall thickness, enhancement of the ascending colon and terminal ileum wall, and enlargement of the abdominal lymph nodes (**Figure 1A**). Intestinal ultrasound showed that the walls of the terminal ileum and colon were thickened. The inflammation affected the vermiform appendix. Colonoscopy was performed, and multiple ulcers in the terminal ileum and colon were detected. The area of the largest ulcer was approximately 3 cm × 4 cm. The ulcers had sharp margins, and most had no exudate or edema; these lesions were mainly located in the ascending colon and cecum. There were also scattered ulcers located in the rectum (**Figure 1B**). Histological study of a mucosal biopsy indicated segmental chronic active inflammation of the colon and terminal ileum, cryptitis and crypt abscess were observed at the colon; additionally, several noncaseating granulomas was observed in the ileum and descending colon (**Figure 1C**). Microbial metagenomic next-generation sequencing (mNGS) of the intestinal biopsy did not reveal any opportunistic pathogens. The patient was diagnosed with VEOIBD and was treated with 2 mg/kg intravenous methylprednisolone twice a day. His fever subsided on the second day of the intravenous methylprednisolone regimen. Then, he was switched to oral prednisone, which was gradually tapered down within 3 months. Repeated colonoscopy showed that the intestinal inflammation was remarkably improved. Only small ulcers in the terminal ileum were observed, and the ulcers located in the colon had healed, leaving a scattering of small inflammatory polyps (**Figure 1D**). Capsule endoscopy confirmed terminal ileal ulcers. The patient took oral mercaptopurine for 4 months. Although he had a wild-type genotype for thiopurine methyltransferase (*TPMT*) and *NUDT15*, he experienced recurrent neutropenia. He was switched to methotrexate (MTX) due to neutropenia, and he had severe sepsis when he was on MTX. He presented with high fever, chill, delirium and shock, accompanied with remarkably elevated CRP and procalcitonin levels. He was treated with vasoactive drug and intravenous antibiotics in the intensive care unit for 1 week. Because of nonspecific intestinal colitis and early onset of disease, he was suspected to have monogenic IBD. The repeated immune work of the peripheral blood was showed in **Supplementary Table S1**. The number and percentage of memory B cells, central $CD8^+$ T cells, Th1 cells and NK cells were obviously reduced.

GENETIC ANALYSIS

The patient and his family underwent whole exome sequencing (WES) and trio analysis due to the early onset of intestinal disease. Peripheral blood samples from the patient and his parents were collected and subjected to next-generation sequencing (NGS) at Running Gene Inc. (Beijing, China). In brief, DNA was isolated from peripheral blood using a DNA Isolation Kit (Blood DNA Kit V2, CW2553). For all samples, shearing worked very consistently, and the peak of the size distribution was approximately 200 bp. DNA libraries were prepared with a KAPA Library Preparation Kit (Kapa Biosystems, KR0453) following the manufacturer's

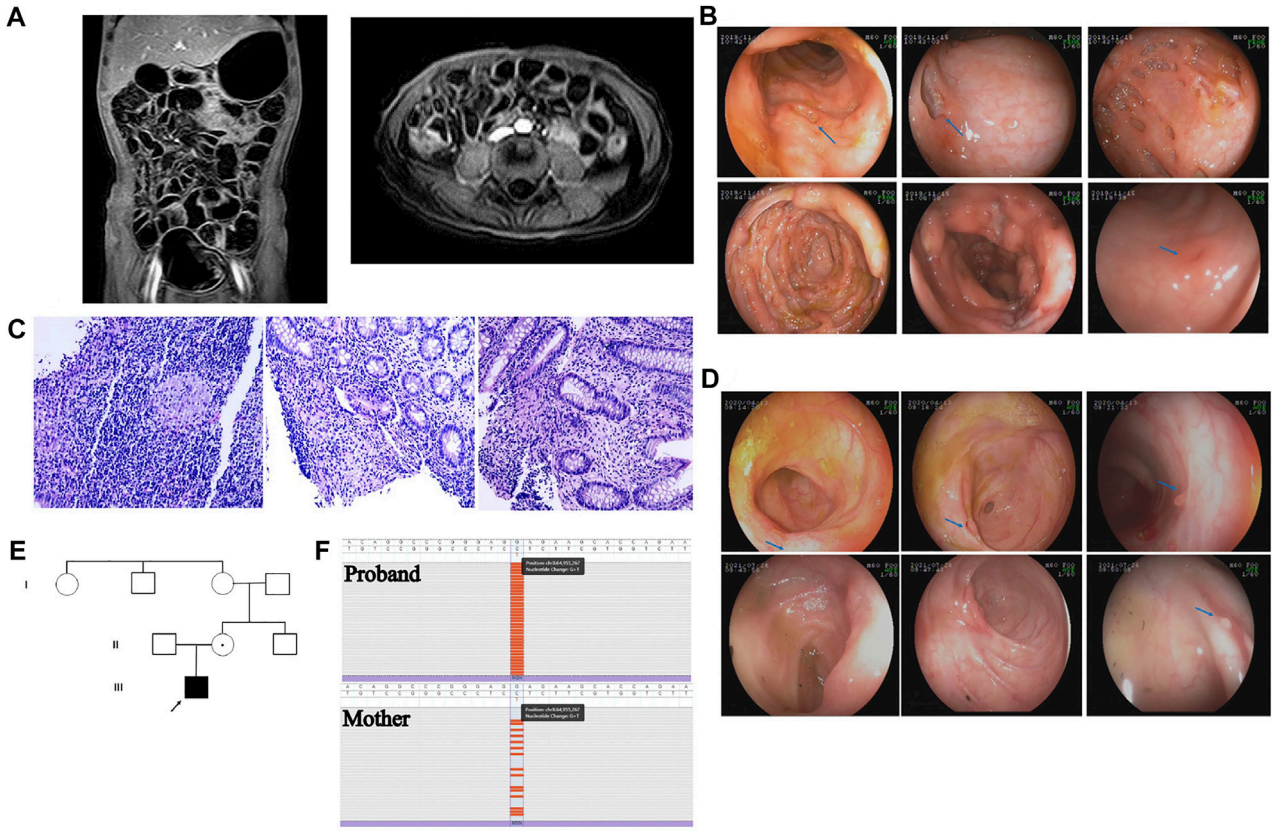


FIGURE 1 | (A) Abdominal CT showed irregular wall thickness and enhancement of ascending colon and terminal ileum wall; enlargement of abdominal lymph nodes **(B)** Colonoscopy detected multiple ulcers in the terminal ileum, colon; and scatter ulcers in the rectum. The ulcers had a sharp margin, and most had no exudate or edema, mainly located at ascending colon and cecum. **(C)** Histological study of the mucosal biopsy indicated segmental chronic active inflammation of the colon and terminal ileum. Cryptitis, crypt abscess and noncaseating granulomas was observed in the ileum and colon **(D)** Repeated colonoscopy showed remarkably improvement of intestinal inflammation. **(E)** A novel nonsense heterozygous mutation of c.934G>T (p.Glu312Ter) was identified in the *MSN* gene. The mutation of the proband was inherited from his mother. **(F)**.

instructions. The libraries were estimated with a Qubit dsDNA HS Assay kit (Invitrogen, Q32851). Array capture, hybridization of pooled libraries to the capture probes and removal of nonhybridized library molecules were carried out according to the SeqCap hybrid mix system. Sequencing, sample dilution, flow cell loading and sequencing were performed according to the Illumina specifications. DNA libraries were sequenced on the Illumina Nova Seq platform as paired-end 200-bp reads. Variant calling and annotation quality control were applied to raw data (stored in FASTQ format) obtained from NovaSeq to guarantee the meaningfulness of downstream analysis. High-quality paired-end reads were aligned to the human reference genome sequence from the UCSC database (build 37.1 version hg19, http://genome.ucsc.edu/) using the Burrows–Wheeler Aligner (https://github.com/lh3/bwa). We estimated quality scores and performed consensus SNP and insertion and deletion (indel) calling using the Genome Analysis Tool Kit. All the called variants were annotated using several public databases (1000 Genomes Project, ExAC, gnomAD, ESP6500, CCDS, RefSeq, Ensembl, etc.). Candidate variants were classified according to the American College of Medical Genetics and Genomics (ACMG) guidelines (Richards et al., 2015).

A novel hemizygous mutation, c.934G>T (p. Glu312Ter), was identified in the *MSN* gene. The pathogenic variant in the proband was inherited from his mother (**Figures 1E,F**). The variant meets PVS1 and PM2, and therefore was likely pathogenic according to ACMG guideline. This pathogenic variant was located in the coiled coil domain of the *MSN*. It was a truncation mutation that resulted in the pre-termination of MSN protein. We specially screened variants that associated with IBD as well as PID, and we did not discover any suspected variant. We hypothesize that this pathogenic variant affects the anchor of the MSN protein and prevents it from playing its role in regulating the actin cytoskeleton in the relevant cells.

DISCUSSION

Variants in the *MSN* gene were recently reported as the cause of a primary immunodeficiency disease called X-MAID by Chantal Lagresle-Peyrou *et al.* in 2016 (Lagresle-Peyrou et al., 2016). They

TABLE 1 | The clinical and genetic mutation features of those reported in literatures along with our case.

	Gene mutation	Mutation derived	Age of VZV infection, years	Age of evaluation, years	Other infections	Skin manifestations and others	Treatment
P1	c.511C>T p.(Arg171Trp)	Mother	3.5	18	Lung infections	Contagiosum molluscum, Eczema	IVIG, SMZ
P2	c.511C>T p.(Arg171Trp)	Mother	1.2	9	Lung infections, bronchitis, gastroenteritis, recurrent severe acute otitis	Contagiosum molluscum, Eczema, alopecia, seborrheic dermatitis	IVIG, SMZ, G-CSF
P3	c.511C>T p.(Arg171Trp)	Mother	1.2	4	Gastroenteritis, *Chryseobacterium* species infection	TTP, cerebrovascular stroke	IVIG, SMZ
P4	c.511C>T p.(Arg171Trp)	NA	1.2	5	NE	Eczema, ecthyma	IVIG, G-CSF, HSCT
P5	c.511C>T p.(Arg171Trp)	Mother	-	3.5	Recurrent gastroenteritis, recurrent upper respiratory tract infections	Eczema, alopecia, Gibert Disease	IVIG, SMZ
P6	c.511C>T p.(Arg171Trp)	NA	5	39	Severe pneumonia with septic shock, recurrent respiratory and urinary tract infections	Contagiosum molluscum, eczema	-
P7	c.1657C>T p.(Arg553Ter))	NA	<10	69	Lung infections bronchitis, recurrent sinopulmonary infections	-	Azithromycin
P8	c.511C>T p.(Arg171Trp)	De novo	-	2.25	Parainfluenzae virus type 2 infection, Oral thrush; enteroditis Rhino/ Entero virus infection	Atopic dermatitis	IVIG
P9	c.511C>T p.(Arg171Trp)	Mother	-	8	Recurrent respiratory infections, diarrhea	Eczema, ecthyma	IVIG, SMZ
P10	c.511C>T p.(Arg171Trp)	Mother	1.5	24	Bronchitis, repeated upper respiratory tract infections, ear infections, periorbital cellulitis, pneumonia	Skin infections, paronychiae, thrombocytopenia	IVIG, G-CSF
P11	c.511 C>T p.(Arg171Trp)	-	-	0.17	Diarrhea, CMV infection	-	SMZ/trimethoprim sulfate, fluconazole, IVIG, HSCT
P12	c.511 C>T p.(Arg171Trp)	-	-	Newborn	-	-	SMZ/trimethoprim sulfate, fluconazole, IVIG, HSCT
Our case	c.934G>T p.(Glu312Ter)	Mother	-	5	Fever, pneumonia, sepsis	Enteropathy	IVIG, steroid, IFX

VZV, varicella-zoster virus; TTP, thrombotic thrombocytopenic purpura; IVIG, intravenous immunoglobin; SMZ, sulfamethoxazole; G-CSF, granulocyte-colony stimulating factor; HSCT, hematopoietic stem cell transplantation; CMV, cytomegalovirus; IFX, infliximab.

investigated 7 male patients from 5 different families who had recurrent bacterial and viral infections, especially varicella-zoster virus (VZV) infections (Lagresle-Peyrou et al., 2016). Subsequently, Ottavia M. Delmonte reported a severe combined immunodeficiency patient who was identified by newborn screening and whose *MSN* mutation was sequenced by WES at 25 months of age (Delmonte et al., 2017). In a 24-year-old man with confirmed immunodeficiency of unknown cause, Gabrielle Bradshaw identified a variant of *MSN* by WES and whole-genome microarray copy number variant (CNV) analysis (Bradshaw et al., 2018). HUI Xiao-ying reported a similar case in a Chinese boy in 2017 (Xiao-chuan et al., 2017). Sarah E. Henrickson et al. (2019) reported three cases of X-MAID who were presented with severe combined immunodeficiency (SCID) and treatment with HSCT. The laboratory features of X-MAID are reduced lymphocytes, low immunoglobin and fluctuating neutropenia (Lagresle-Peyrou et al., 2016). There have been only five articles on this disease worldwide, describing a total of 12 patients from nine families; furthermore, there is a ESID registry study calling for an international, multicenter, retrospective study of patients with X-MAID, and they have collected about a total of 16 patients. Here, we summarized the clinical and genetic mutation features of reported X-MAID patients along with those of our case in **Table 1**. All the patients were male, and the age of genetic diagnosis ranged from newborn to 69 years, with median age of 5 years. The clinical manifestations were recurrent bacterial and viral infections of the respiratory and gastrointestinal systems. Eight patients had skin manifestations, mainly including eczema and molluscum contagiosum. Common immune-associated laboratory features of reported X-MAID are summarized in **Table 2**, including low neutrocyte levels (11/13), low levels of lymphocytes (12/13), reduced $CD19^+$ B lymphocytes (13/13), reduced $CD4^+$ (13/13) and $CD8^+$ T cell counts (12/13), and low levels of IgG (8/10). While our patient presented a different phenotype. He did not have obvious recurrent infections till now, and the predominant clinical feature was IBD-like intestinal inflammatory.

Although *MSN* mutation affected both T cell and B cell function, patients with X-MAID presented with deficiencies in cellular immunity and humoral immunity. These patients had a relatively good prognosis. The oldest patient was 69 years old, and the most severe condition was severe pneumonia accompanied by

TABLE 2 | The immune associate results of reported X-MAID patients and our case.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	This case
Neutrocytes	L	L	L	L	L	L	N	L	L	N	L	L	L
Monocytes	L	L	L	N	L	L	L	L	L	N	ND	ND	L
Lymphocytes	L	L	L	L	L	L	L	L	L	L	L	L	N
$CD19^+$ cells	L	L	L	L	L	L	L	L	L	L	L	L	L
Naive B cells	L	L	L	ND	ND	ND	ND	ND	N	ND	ND	ND	L
Natural killer cells	L	L	L	L	L	L	L	ND	ND	L	N	L	L
$CD4^+$ cells	L	L	L	L	L	L	L	L	L	L	L	L	L
$CD8^+$cells	L	L	L	L	L	N	L	L	L	L	L	L	L
Naive $CD4^+$T cells	L	L	L	ND	L	L	L	N	L	ND	ND	ND	N
Naive $CD8^+$T cells	L	L	L	ND	L	L	L	L	N	ND	ND	ND	N
$CD57^+CD8^+$ T cells	H	H	H	ND	H	H	H	ND	ND	ND	ND	ND	N
IgG	L	L	L	L	NE	L	L	L	L	N	ND	ND	N
IgA	L	N	L	L	L	L	N	N	L	L	ND	ND	N
IgM	L	L	L	L	L	L	L	L	L	L	ND	ND	N
IgE	NE	N	NE	N	N	NE	NE	N	N	ND	ND	ND	N

L, low; H, high; N, normal; NE, Non-evaluated; ND, not determined; Ig, immunoglobin.

shock. Other complications were common infections and were treated with immunoglobin infusion and antibiotics. Sulfamethoxazole (SMZ) tablets and recombinant human granulocyte colony-stimulating factor (G-CSF) were used to prevent opportunistic infections for recurrent neutropenia. Three patients exhibited severe combined immunodeficiency received hematopoietic stem cell transplantation (HSCT). However, mutations in the *MSN* gene that can lead to IBD-like feature have not been reported, and the optimal treatment of colitis caused by *MSN* variant is not clear. Remission of inflammatory colitis was induced with steroid and maintained with anti-TNF therapy in this case. There is no standard treatment for VEOIBD or IBD-like features caused by immunodeficiencies, the main treatment included supporting treatment, symptomatic treatment and conventional treatment for IBD, such as steroids, immunosuppressants, biological treatment and thalidomide, et al. The patient in our case had persistent fever, no chronic history of disease, and the endoscopy and pathological finding of biopsy indicated the feature of IBD, so we used steroid to induce remission. The patient did not have allergy history, and his nutrition status was normal, so he did not have any special diet.

The human moesin gene, *MSN,* is located on chromosome Xq12 and has 13 exons distributed over more than 30 kb. Only two different variants of the *MSN* gene have been reported. Eleven patients were identified to harbor the same variant of c.511C>T (p. Arg171Trp), which is located in the FERM domain. Another reported variant is c.1657C>T (p.Arg553Ter), which is located in the F-actin binding domain. Our case has a nonsense variant of c.934G>T (p.Glu312Ter) located in the coiled coil domain. This variant alters the codon for Glu312 to a termination codon, which causes premature termination of MSN protein synthesis. Our case is the second reported nonsense variant in *MSN* gene, and the phenotype is differed from other pathogenic variants, which may suggest that nonsense variants in *MSN* may cause IBD-like features. The variant of c.934G>T (p.Glu312Ter) is classified as a variant of likely pathogenic according to the ACMG guidelines (Richards et al., 2015). The pLi score for *MSN* in gnomAD is 1, which means the gene is highly intolerant of loss of function variants (Ziegler et al., 2019), this adds further evidence that *MSN* variants cause immunodeficiency and widen the reported phenotypes.

MSN protein is mainly expressed in endothelial cells, lymphocytes and platelets, and it regulates the actin cytoskeleton. *MSN* knockout mice was observed to impair the migration and adhesion of lymphocyte cells; naive $CD8^+$ T cells were affected most. Recently, isolated $CD8^+$ Treg cells from moesin-deficient mice exhibited impaired proliferation in response to IL-15, which shows the importance of moesin in IL-15–dependent $CD8^+$ Treg cell homeostasis (Satooka et al., 2017). Additionally, moesin regulates NK cell survival through IL-15–mediated signaling *in vivo* (Satooka et al., 2021). Altogether, *MSN* mutation may affect both innate and adaptive immunity *in vivo*. However, the underlying mechanism by which variants of *MSN* causes IBD-like features needs further investigation.

PIDs have a cross-connection with VEOIBD. The actin cytoskeleton is required for many immune cell functions, including migration, adhesion, phagocytosis, assembly of complex intercellular contacts, and cell division (Burns et al., 2017). PID-associated actin regulatory defects include Wiskott-Aldrich syndrome (WAS), WASp-interacting protein (WIP) deficiency, *DOCK2* and *DOCK8* deficiencies, moesin deficiency, leukocyte adhesion defects (LAD1, 2, 3 and IV) and megakaryoblastic leukemia 1 (*MKL1*), and these defects lead to combined immunodeficiencies or phagocyte disorders (Burns et al., 2017). It has been reported that WAS can present as IBD-like intestinal inflammation (Ohya et al., 2017). We suspect that the mechanism of *MSN*-associated colitis is as follows. First, similar to WAS, MSN is a member of the cell-structure-related proteins that regulates the cell actin cytoskeleton and plasma membrane. The actin cytoskeleton is required for many immune cell functions. Dysfunction of immune cells is related to nonspecific colitis or IBD-like features. Secondly,

MSN is dominantly expressed on lymphocytes, which plays an important role in the regulation of intestinal inflammation. Chronic intestinal inflammation could lead to IBD-like features. In our case, the boy exhibited severe intestinal inflammation at first, a decrease in B lymphocytes and neutrocytes, low albumin, and elevation of CRP and ESR, and he was not responsive to intravenous antibiotics treatment. The negative result of blood serologic tests, immunohistochemistry study and the mNGS test of the mucosal biopsy did not indicate any infection that caused the inflammation of the colon. The WES did not reveal any causative variant that related to PID or VEOIBD. After the administration of steroid, the inflammatory markers and intestinal inflammation were obviously improved, which indicated that an immune mechanism participated in the process of inflammation. However, other typical phenotypes of X-MAID patients, such as hypogammaglobulinemia, neutropenia, VAZ infection, recurrent respiratory tract infections, and eczema, were not obvious in this patient. The mechanism of intestinal inflammation in this patient is not clear, but we presume that it may be associated with dysregulation of immune cells and altered functioning of lymphocytes. Intriguingly, some immunocompromising diseases, such as common variable immunodeficiency, hyper-IgM syndrome, and selective IgA deficiency, present with IBD-like intestinal inflammation (Albshesh et al., 2021). In an immunocompromised patient, opportunistic infections, such as cytomegalovirus or Epstein-Barr virus infection, may contribute to intestinal inflammation as well. However, this patient did not have recurrent infections, and his pathogen tests were negative. The mNGS of an intestinal biopsy was negative as well. The patient's colon ulcers were almost completely healed after steroid treatment, which may indicate that the ulcers were induced by immune dysregulation. He could not tolerate mercaptopurine or MTX, so he was ultimately switched to IFX to induce and maintain mucosal healing. The intestinal ulcers were healed after three doses of infused IFX.

In conclusion, we reported a patient with nonsense variant of *MSN* who presented with IBD-like features. This is the first reported case of an X-MAID patient maily presented with enteropathy. The pathogenesis of intestinal colitis is not clear and requires further basic medical research.

DATA AVAILABILITY STATEMENT

The datasets for this article are not publicly available due to concerns regarding participant/patient anonymity. Requests to access the datasets should be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Children's Hospital of Zhejiang University School of Medicine. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the patient's parents for the investigation and publication of this article, including all data and images.

AUTHOR CONTRIBUTIONS

YF wrote the manuscript; YL collected the clinical data; YL interpreted the pathological result. JC designed the study.

FUNDING

Supported by the Natural Science Foundation of Zhejiang Province (LQ19H030010).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.873635/full#supplementary-material

Supplementary Table S1 | The number and percentage of peripheral blood immune cells of the patient.

REFERENCES

Albshesh, A., Eder, P., Ribaldone, D. G., Oldenburg, B., de Boer, N. K., Mantzaris, G. J., et al. (2021). Primary Hypogammaglobulinaemia with Inflammatory Bowel Disease-like Features: An ECCO CONFER Multicentre Case Series. *J. Crohns Colitis* 16, 91–97. doi:10.1093/ecco-jcc/jjab124

Berryman, M., Franck, Z., and Bretscher, A. (1993). Ezrin Is Concentrated in the Apical Microvilli of a Wide Variety of Epithelial Cells whereas Moesin Is Found Primarily in Endothelial Cells. *J. Cell. Sci.* 105, 1025–1043. doi:10.1242/jcs.105.4.1025

Bradshaw, G., Lualhati, R. R., Albury, C. L., Maksemous, N., Roos-Araujo, D., Smith, R. A., et al. (2018). Exome Sequencing Diagnoses X-Linked Moesin-Associated Immunodeficiency in a Primary Immunodeficiency Case. *Front. Immunol.* 9, 420. doi:10.3389/fimmu.2018.00420

Burns, S. O., Zarafov, A., and Thrasher, A. J. (2017). Primary Immunodeficiencies Due to Abnormalities of the Actin Cytoskeleton. *Curr. Opin. Hematol.* 24 (1), 16–22. doi:10.1097/MOH.0000000000000296

Charbit-Henrion, F., Parlato, M., Hanein, S., Duclaux-Loras, R., Nowak, J., Begue, B., et al. (2018). Diagnostic Yield of Next-Generation Sequencing in Very Early-Onset Inflammatory Bowel Diseases: A Multicentre Study. *J. Crohns Colitis* 12 (9), 1104–1112. doi:10.1093/ecco-jcc/jjy068

Delmonte, O. M., Biggs, C. M., Hayward, A., Comeau, A. M., Kuehn, H. S., Rosenzweig, S. D., et al. (2017). First Case of X-Linked Moesin Deficiency Identified after Newborn Screening for SCID. *J. Clin. Immunol.* 37 (4), 336–338. doi:10.1007/s10875-017-0391-9

Henrickson, S. E., Andre-Schmutz, I., Lagresle-Peyrou, C., Deardorff, M. A., Jyonouchi, H., Neven, B., et al. (2019). Hematopoietic Stem Cell Transplant for the Treatment of X-MAID. *Front. Pediatr.* 7, 170. doi:10.3389/fped.2019.00170

Hirata, T., Nomachi, A., Tohya, K., Miyasaka, M., Tsukita, S., Watanabe, T., et al. (2012). Moesin-deficient Mice Reveal a Non-redundant Role for Moesin in Lymphocyte Homeostasis. *Int. Immunol.* 24 (11), 705–717. doi:10.1093/intimm/dxs077

Kelsen, J. R., Sullivan, K. E., Rabizadeh, S., Singh, N., Snapper, S., Elkadri, A., et al. (2020). North American Society for Pediatric Gastroenterology, Hepatology, and Nutrition Position Paper on the Evaluation and

Management for Patients with Very Early-Onset Inflammatory Bowel Disease. *J. Pediatr. Gastroenterol. Nutr.* 70 (3), 389–403. doi:10.1097/MPG.0000000000002567

Lagresle-Peyrou, C., Luce, S., Ouchani, F., Soheili, T. S., Sadek, H., Chouteau, M., et al. (2016). X-linked Primary Immunodeficiency Associated with Hemizygous Mutations in the Moesin (MSN) Gene. *J. Allergy Clin. Immunol.* 138 (6), 1681–1689. doi:10.1016/j.jaci.2016.04.032

Ohya, T., Yanagimachi, M., Iwasawa, K., Umetsu, S., Sogo, T., Inui, A., et al. (2017). Childhood-onset Inflammatory Bowel Diseases Associated with Mutation of Wiskott-Aldrich Syndrome Protein Gene. *Wjg* 23 (48), 8544–8552. doi:10.3748/wjg.v23.i48.8544

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (2015). Standards and Guidelines for the Interpretation of Sequence Variants: a Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* 17 (5), 405–424. doi:10.1038/gim.2015.30

Satooka, H., Matsui, M., Ichioka, S., Nakamura, Y., and Hirata, T. (2021). The ERM Protein Moesin Regulates Natural Killer Cell Homeostasis *in vivo*. *Cell Immunol.* 371, 104456. doi:10.1016/j.cellimm.2021.104456

Satooka, H., Nagakubo, D., Sato, T., and Hirata, T. (2017). The ERM Protein Moesin Regulates CD8(+) Regulatory T Cell Homeostasis and Self-Tolerance. *J Immunol.* 199 (10), 3416–3426. doi:10.4049/jimmunol.1700074

Xiao-ying, H., Jin-qiao, S., Wen-jie, W., Ying, W., Bing-bing, W., and Xiao-chuan, W. (2017). Primary Immunodeficiency Disease Caused by MSN Gene Mutation in One Child and Literature Review. *Chin. J. Evid. Based Pediatr.* 12 (4), 300–303. doi:10.3969/j.issn.1673-5501.2017.04.012

Ziegler, A., Colin, E., Goudenège, D., and Bonneau, D. (2019). A Snapshot of Some pLI Score Pitfalls. *Hum. Mutat.* 40 (7), 839–841. doi:10.1002/humu.23763

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.

Copyright © 2022 Fang, Luo, Liu and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 13 June 2022
doi: 10.3389/fgene.2022.941961

Causal Effects of Circulating Cytokines on the Risk of Psoriasis Vulgaris: A Mendelian Randomization Study

Pan Zhao, Jing Zhang, Biyong Liu, Yufei Tang, Lei Wang, Guifeng Wang, Huihui Wu, Chengwei Yang, Xuemei Li and Bo Li*

Anhui Prevention and Treatment Center for Occupational Disease, Anhui No. 2 Provincial People's Hospital, Hefei, China

Background: Psoriasis vulgaris is an inflammatory skin disease. Observational studies have shown associations between circulating cytokine levels and psoriasis vulgaris. But the causal relationship between circulating cytokine and psoriasis vulgaris remains elusive.

Methods: To assess the causal effects of cytokine levels on the risk of psoriasis vulgaris and vice versa, we performed a two-sample Mendelian randomization (MR) study by using the inverse-variance weighted (IVW), weighted median, and Mendelian randomization pleiotropy residual sum and outlier (MR-PRESSO) in genome-wide association summary statistics of 41 circulating cytokines in up to 8,293 individuals and psoriasis vulgaris in 399,883 individuals.

Results: We identified that increasing RANTES level induced an elevated risk of psoriasis vulgaris in IVW (β = 0.33, S.E. = 0.12, p = 0.006). This causal effect showed consistency across the weighted median (β = 0.35, S.E. = 0.15, p = 0.022) and MR-PRESSO method (β = 0.33, S.E. = 0.11, p = 0.028).

Conclusions: Our results suggest a potential causal effect of elevated RANTES concentration on the increased risk of psoriasis vulgaris.

Keywords: psoriasis vulgaris, cytokines, causal effect, mendelian randomization, RANTES

OPEN ACCESS

Edited by:
Xianyong Yin,
University of Michigan, United States

Reviewed by:
Changbing Shen,
Peking University, China
Fusheng Zhou,
Anhui Medical University, China

****Correspondence:***
Pan Zhao
zhaopanmj@163.com

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: *12 May 2022*
Accepted: *24 May 2022*
Published: *13 June 2022*

Citation:
Zhao P, Zhang J, Liu B, Tang Y, Wang L, Wang G, Wu H, Yang C, Li X and Li B (2022) Causal Effects of Circulating Cytokines on the Risk of Psoriasis Vulgaris: A Mendelian Randomization Study. Front. Genet. 13:941961. doi: 10.3389/fgene.2022.941961

INTRODUCTION

Psoriasis vulgaris is a chronic inflammatory skin disease, affecting about 2% of the population around the world (von Csiky-Sessoms and Lebwohl, 2019; Rendon and Schäkel, 2019). The prevalence of psoriasis vulgaris varies across regions, ranging from 0.05 to 0.47% in Asia to 2–3% in Europe and the United States (Hawkes et al., 2017; Kaufman and Alexis, 2018). It is clinically characterized by erythema lesions covered with pearly scales (Hawkes et al., 2017; Rendon and Schäkel, 2019). The exact cause of psoriasis vulgaris is not well understood. Observational studies have shown associations between psoriasis vulgaris and circulating cytokines such as IL-23, IL-17 and TNF-α (de Alcantara et al., 2021). However, the causal relationship between cytokines and psoriasis vulgaris remain unresolved.

Since observational studies are susceptible to reverse causality and confounding factors, it is unable to provide convincing evidence for causality alone (Smith and Ebrahim, 2004). For instance, elevated IL-22 levels have been observed in both skin lesions and blood, the strategy of blocking IL-22

in the treatment of psoriasis failed in a clinical trial (Chiricozzi et al., 2018). Studies show that plasmacytoid dendritic cells in psoriatic lesions produce IFN-α, suggesting a key role of IFN-α in the pathogenesis of psoriasis (Gilliet et al., 2004; Nestle et al., 2005). However, a randomized trial failed to provide evidence that targeting IFN-α improved clinical symptoms in patients with plaque psoriasis (Bissonnette et al., 2010). Many studies detect an increasing level of IL-17 in the occurrence and development of psoriasis (Harper et al., 2009; Johansen et al., 2009; Ramirez-Carrozzi et al., 2011). But in an experimental study, IL-17 restored the function of keratinocytes and played a protective role in the development of psoriasis (Kanda et al., 2005; Deng et al., 2016).

Mendelian randomization (MR) uses genetic variation that are strongly associated with a possible exposure as instrumental variables (IVs) to assess the causality of the given exposure on an outcome of interest (Emdin, et al., 2017). Compared with observational studies, MR is immune from reverse causality and confounding factors. In past decade, genome-wide association studies (GWAS) have identified tens of thousands of genetic variants robustly associated with complex traits (Hemani, et al., 2018). These findings reveal the genetic underpinning of complex traits. The associated genetic variant can be used as IVs and facilitate the application of MR to identify disease's causal factors (Davey Smith and Hemani, 2014).

To better understand the causal role of cytokines in the development of psoriasis vulgaris and to search for disease prevention and treatment targets of potential, we performed a two-sample MR to evaluate the causal relationship between circulating cytokines concentrations and the risk of psoriasis vulgaris. We identified single nucleotide polymorphisms (SNPs) from GWAS meta-analysis for 41 circulating cytokines in 8,293 Finns as IVs. We evaluated the causal effect of circulating cytokines on psoriasis vulgaris with SNPs which were identified, and the genome-wide association statistics for psoriasis vulgaris in the United Kingdom biobank. Our study suggests a potential causal association between circulating RANTES levels and the risk of psoriasis vulgaris. This finding helps identify a potential target for the intervention of psoriasis.

MATERIALS AND METHODS

Genome-Wide Association Summary Statistics

We accessed genome-wide association summary-level statistics for 41 circulating cytokines and psoriasis vulgaris. In brief, the summary-level data for 41 cytokines included ~ 10.7 million SNPs and 8,293 Finnish individuals from the Cardiovascular Risk in Young Finns study, FINRISK 1997, and FINRISK 2002 (Ahola-Olli et al., 2017). The cytokine levels were quantified by Bio-Rad's premixed Bio-Plex Pro Human Cytokine 27-plex Assay and 21-plex Assay, and Bio-Plex 200 reader with Bio-Plex 6.0 software. Genotype imputation inference was based on reference haplotypes from the 1000 Genomes Project Phase 1 (1000 Genomes Project Consortium et al., 2010). The single-variant association was accomplished through a linear regression between cytokine levels and SNPs after adjusting for age, sex, and body mass index (Ahola-Olli et al., 2017).

We used genome-wide association statistics for psoriasis vulgaris in the United Kingdom biobank (Sudlow et al., 2015). The United Kingdom Biobank defined psoriasis vulgaris using International Classification of Diseases (ICD) L40.0 based on the medical records and questionnaire data for each participate. The United Kingdom biobank data contains 1,684 psoriasis vulgaris cases and 398,199 controls of White British. GWAS was accomplished between the risk of psoriasis vulgaris and 31.3 million SNPs through SAIGE (Sudlow et al., 2015; Zhou et al., 2018). About 8.4 million SNPs were overlapped with the GWAS for circulating cytokines. No ethical approval was required.

Identification of IVs

The IVs for MR have three assumptions: 1) IVs must be strongly associated with exposure. 2) IVs are not associated with any known confounders related to outcome or exposure. 3) IVs cannot affect the outcome in other ways except through exposure (Lawlor et al., 2008). To identify IVs for MR analysis, we performed linkage disequilibrium (LD) pruning in Plink 2.0 in the 8.4 million overlapping SNPs at $P < 1 \times 10^{-6}$ in cytokine GWAS summary statistics using LD $r^2 < 0.1$ in distance≥500 kilobases (kb) (Machiela and Chanock, 2015; Sobota et al., 2015). After screening, 420 SNPs associated with 40 cytokines were used for further analyses as IVs. To attain a stable assessment, we restricted our MR analyses to 35 of the 41 cytokines with the number of IVs ≥3.

MR Analysis

To test the potential causal effects of circulating cytokine levels on the risk of psoriasis vulgaris and vice versa, we ran MR analysis by using three two-sample MR methods: IVW, weighted median, and MR-PRESSO (Burgess et al., 2017; Verbanck et al., 2018). These three methods make different assumptions and deal with the potential horizontal pleiotropy differently. The IVW method integrates causal effects assessment from multiple genetic variants, assuming that all IVs are valid and free of interference from heterogeneity and pleiotropy (Pierce and Burgess, 2013). When at least 50% SNPs are valid instruments, weighted median test is recommended (Bowden et al., 2016). The MR-PRESSO test is appropriate when less than 50% of the genetic variants have a pleiotropic effect (Verbanck et al., 2018). We required the causal effects consistent across all the three methods and used a significance threshold $p < 0.05$ to define a significant causal effect (Burgess et al., 2017; Verbanck et al., 2018).

Test for Heterogeneity and Pleiotropy

To evaluate the validity of the MR core hypothesis, we performed a series of assessments of heterogeneity and pleiotropy. MR-PRESSO method can test the directional pleiotropy and identify and correct potential outliers (Verbanck et al., 2018). We evaluated horizontal pleiotropy by examining the intercept term of MR Egger's regression (Bowden et al., 2015). If the intercept term is significantly different from zero at $p < 0.05$, horizontal pleiotropy exists (Bowden et al., 2015). We calculated a Cochran's Q-statistic to quantify the heterogeneity of causal effects across all IVs (Egger et al., 1997).

TABLE 1 | MR analysis of the causal relationship between RANTES, SDF-1α, MIP-1β and IL-17 levels and psoriasis vulgaris.

Cytokines	No. of SNPs	Association			Heterogeneity	MR-PRESSO global Test
		β	S.E.	*P*-Value	*P*-Value	*P*-Value
RANTES						
IVW	6	0.325	0.119	**0.006**	0.55	—
Weighted median	6	0.348	0.153	**0.022**	—	—
MR-PRESSO	6	0.325	0.106	**0.028**	—	0.617
SDF-1α						
IVW	3	–0.533	0.221	**0.016**	0.37	—
Weighted median	3	–0.516	0.267	0.053	—	—
MR-PRESSO	3	—	—	—	—	—
MIP-1β						
IVW	145	–0.047	0.030	0.117	0.02	—
Weighted median	145	–0.096	0.044	**0.028**	—	—
MR-PRESSO	145	NA	NA	NA	—	0.023
IL-17						
IVW	4	0.310	0.381	0.416	0.02	—
Weighted median	4	0.512	0.295	0.082	—	—
MR-PRESSO	4	0.623	0.105	**0.027**	—	0.027

Note: Significant results are shown in bold.

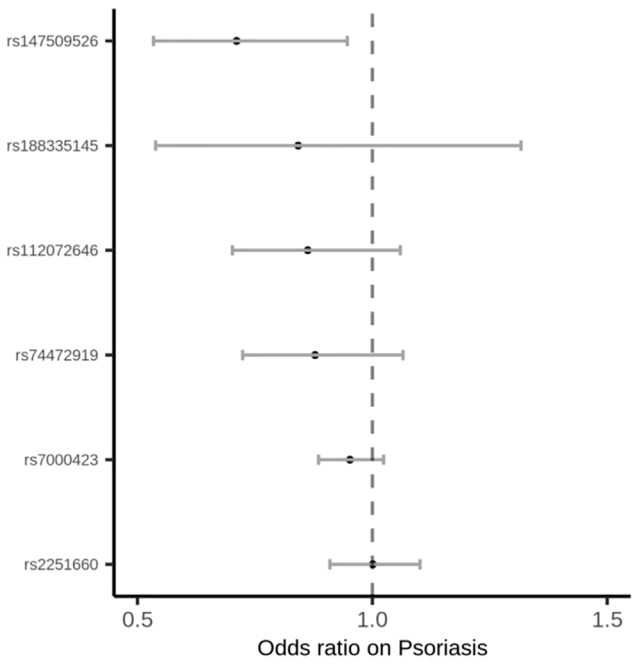


FIGURE 1 | Forest plot of the effects of IVs for RANTES on the risk of psoriasis vulgaris.

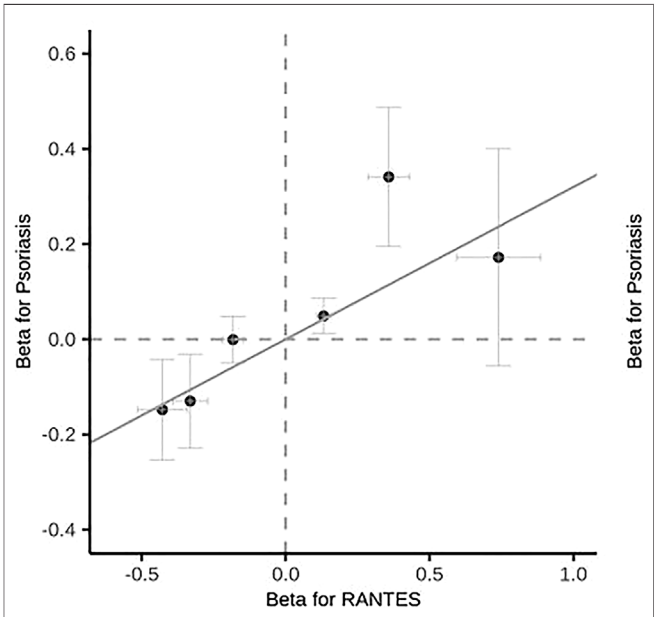


FIGURE 2 | Scatter plot of the effects of IVs on RANTES levels and the risk of psoriasis vulgaris.

RESULTS

Selection of IVs

For 35 of the 41 cytokines, we identified at least three genetic variants available at $p < 1 \times 10^{-6}$ and LD ($r^2 < 0.1$, distance≥500 kb). The numbers of IVs for each cytokine are shown in **Table 1**. The associations of IVs with each circulating cytokine level can be found in **Supplementary Table S1**.

Causal Effects of Cytokine Levels on the Risk of Psoriasis Vulgaris

Among the 35 cytokines, genetically predicted RANTES, SDF-1α, MIP-1β and IL-17 were associated with the risk of psoriasis vulgaris in at least 1 MR method at the significance threshold ($p < 0.05$; **Table 1**). MR results of the rest 31 cytokines in our analyses can be found in **Supplementary Table S2**. In the four cytokines RANTES, SDF-1α, MIP-1β, and IL-17, only the causal effect of RANTES on psoriasis vulgaris was significant in all the 3 MR methods: IVW ($\beta = 0.325$, S.E. = 0.119, $p = 0.006$), weighted median ($\beta = 0.348$, S.E. = 0.153, $p = 0.022$), and MR-PRESSO ($\beta =$

0.325, S.E. = 0.106, p = 0.028; **Table 1** and **Figures 1**, **2**). We identified six IVs for the MR analysis evaluating the causal effect of RANTES. The causal effects of these six IVs on the risk of psoriasis vulgaris showed no statistically significant heterogeneity (Cochran's Q-test p = 0.55; **Table 1**). We observed no evidence of horizontal pleiotropy in the MR-PRESSO (Global test p = 0.62; **Table 1**) or the MR-Egger regression (p for intercept = 0.69; **Supplementary Table S3**).

Causal Effects of Psoriasis Vulgaris on Circulating Cytokine Levels

Reversely, we identified causal effects of the onset of psoriasis vulgaris on 11 circulating cytokine levels including IL-1ra, IL-5, CTACK, MIG, VEGF, IL-8, IL-9, IL-13, IL-12p70, TNF-α, and IL-2 at $p < 0.05$ in at least one of the 3 MR methods based on MR results from **Supplementary Table S4**. Out of these 11 cytokines, the causal effects of psoriasis vulgaris achieved consistently significant on IL-1ra, IL-5, and MIG in all the 3 MR methods. We did not detect significant heterogeneity of causal effects (Cochran's Q-test $p > 0.05$; **Supplementary Table S4**) or horizontal pleiotropy (MR-PRESSO Global test $p > 0.05$ and MR Egger's regression intercept $p > 0.05$; **Supplementary Tables S4, S5**).

DISCUSSION

Studies have shown significant associations between circulating cytokines and psoriasis vulgaris, but the causal relationship between cytokines and psoriasis vulgaris remain unclear. In our study, we leveraged large-scale GWAS for 41 circulating cytokines and psoriasis vulgaris and evaluated the causal relationship between circulating cytokines and psoriasis vulgaris by using three two-sample MR approaches. We identified a causal effect of elevated RANTES levels on the increased risk of psoriasis vulgaris. We also assessed the causal effects of psoriasis vulgaris risk on circulating cytokine concentrations, which found that increased risk of psoriasis vulgaris resulted in elevated circulating levels of IL-1ra, IL-5, and MIG.

Psoriasis vulgaris is characterized by abnormal epidermal hyperplasia and differentiation, with accumulation and infiltration of neutrophils and pro-inflammatory T cells in the *epidermis* and dermis (Chiricozzi et al., 2018). RANTES is a pro-inflammatory chemokine, which not only mediates the infiltration of lymphocytes, granulocytes, natural killer cells, and dendritic cells into the inflammatory lesions, but also activates leukocytes to play a role in the inflammatory response (Appay and Rowland-Jones, 2001; Levy, 2009; Marques et al., 2013). RANTES has been reported to play an important role in leukocyte recruitment to psoriatic skin lesions (Nedoszytko et al., 2014; Zablotna et al., 2016). Studies have shown that RANTES production increased in psoriatic lesions compared with non-lesion skin, and was mainly present in keratinocytes or the intercellular space between keratinocytes in psoriatic lesions, from the middle to the marginal lesions of psoriatic plaques (Rateb et al., 2012; Johansen et al., 2017; Joshi et al., 2019). The expression of RANTES's receptor CCR5 was enhanced in epidermal T cells, and the number of CCR5 positive cells raised significantly too (de Groot et al., 2007; Joshi et al., 2019). Phototherapy has been reported to inhibit the production of RANTES in patients with psoriasis. After treatment with NB-UVB, the expression of RANTES was significantly reduced (Rateb et al., 2012; Joshi et al., 2019). According to several case-control trials, serum concentrations of RANTES were significantly higher in patients with psoriasis compared with healthy controls (Rateb et al., 2012; Duarte et al., 2015; Zablotna et al., 2016). However, in a recent case-control study from northern India, serum RANTES levels were significantly reduced in 40 patients with chronic plaque psoriasis compared with 25 healthy controls (Joshi et al., 2019). Previous studies on the relationship between RANTES and psoriasis were observational, difficult to exclude potential confounding factors, and the sample size was small. In this study, we performed a two-sample MR using large sample size GWAS data and found that high RANTES expression was associated with increased risk of psoriasis vulgaris, indicating the positive role of the genetically predicted RANTES for psoriasis vulgaris risk.

We consistently found that psoriasis vulgaris affects the circulating levels of IL-1ra, IL-5, and MIG, which are consistent with previous findings. Abundant IL-ra expression was found in psoriatic lesions (Hammerberg et al., 1992). IL-ra expression was elevated in peripheral blood mononuclear cells of psoriasis patients compared with controls (Kim et al., 2016), while serum IL-5 levels were significantly elevated in psoriatic patients and were significantly overexpressed during the active phase of psoriasis vulgaris (Dong et al., 2021). MIG is induced by IFN-γ and can induce T cells by chemotaxis (Duarte et al., 2015). Studies showed that MIG was highly expressed in psoriatic patients and mediated T cell aggregation and activation (Goebeler et al., 1998; Duarte et al., 2015).

Compared with traditional observational studies, our MR study overcomes the bias caused by confounding and reverse causality issues and provides more reliable evidence for assessing the causal relationship between circulating cytokine concentrations and the risk of psoriasis vulgaris. However, there are some limitations in our MR study, which may influence the causal inference between cytokines and psoriasis vulgaris. Firstly, the sample size of the cytokine GWAS was limited, limiting the statistical power of causal assessment in MR. This may explain IL-17, IL-23, and TNF-α whose biologics are effective for treating psoriasis but did not yield statistically significant results in our MR analysis. Secondly, we used the significance p-value threshold at 1×10^{-6} instead of 5×10^{-8}. The lenient P-value threshold increased the number of available IVs for MR but might cause weak IV bias. Thirdly, our MR study could only demonstrate the lifetime effects of higher cytokines on psoriasis vulgaris, and the effect of short-term regulation of cytokines, such as biological agents targeting cytokines, on psoriasis vulgaris is difficult to assess. Therefore, in the future studies, we need to further explore the biological functions of

cytokines to better understand the relationship between cytokines and psoriasis vulgaris.

CONCLUSION

We identified a causal effect of circulating RANTES levels on the risk of psoriasis vulgaris. This finding might help explore the mechanisms of RANTES in the pathogenesis and development of psoriasis and guide the evaluating of RANTES as a potential intervention target for psoriasis vulgaris.

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

PZ oversaw the conception of the study, design, as well as the collection, analysis and interpretation of data, and drew up the manuscript. JZ and BL were involved in the design of the research. YT, LW, and GW undertook some data analysis and interpretation work. HW, CY, XL, and BL contributed to data collection and interpretation.

FUNDING

Our research work was supported by Anhui Provincial Natural Science Foundation (2008085MH261), and Science Foundation of Anhui Medical University (2021xkj119).

ACKNOWLEDGMENTS

We want to acknowledge the United Kingdom biobank for supporting our work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.941961/full#supplementary-material

REFERENCES

1000 Genomes Project Consortium Abecasis, G. R., Altshuler, D., Auton, A., Brooks, L. D., Durbin, R. M., Gibbs, R. A., et al. (2010). A Map of Human Genome Variation from Population-Scale Sequencing. *Nature* 467 (7319), 1061–1073. doi:10.1038/nature09534

Ahola-Olli, A. V., Würtz, P., Havulinna, A. S., Aalto, K., Pitkänen, N., Lehtimäki, T., et al. (2017). Genome-wide Association Study Identifies 27 Loci Influencing Concentrations of Circulating Cytokines and Growth Factors. *Am. J. Hum. Genet.* 100 (1), 40–50. doi:10.1016/j.ajhg.2016.11.007

Appay, V., and Rowland-Jones, S. L. (2001). RANTES: a Versatile and Controversial Chemokine. *Trends Immunol.* 22 (2), 83–87. doi:10.1016/s1471-4906(00)01812-3

Bissonnette, R., Papp, K., Maari, C., Yao, Y., Robbie, G., White, W. I., et al. (2010). A Randomized, Double-Blind, Placebo-Controlled, Phase I Study of MEDI-545, an Anti-interferon-alfa Monoclonal Antibody, in Subjects with Chronic Psoriasis. *J. Am. Acad. Dermatology* 62 (3), 427–436. doi:10.1016/j.jaad.2009.05.042

Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian Randomization with Invalid Instruments: Effect Estimation and Bias Detection through Egger Regression. *Int. J. Epidemiol.* 44 (2), 512–525. doi:10.1093/ije/dyv080

Bowden, J., Davey Smith, G., Haycock, P. C., and Burgess, S. (2016). Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet. Epidemiol.* 40 (4), 304–314. doi:10.1002/gepi.21965

Burgess, S., Bowden, J., Fall, T., Ingelsson, E., and Thompson, S. G. (2017). Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. *Epidemiology* 28 (1), 30–42. doi:10.1097/EDE.0000000000000559

Chiricozzi, A., Romanelli, P., Volpe, E., Borsellino, G., and Romanelli, M. (2018). Scanning the Immunopathogenesis of Psoriasis. *Int. J. Mol. Sci.* 19 (1), 179. doi:10.3390/ijms19010179

Davey Smith, G., and Hemani, G. (2014). Mendelian Randomization: Genetic Anchors for Causal Inference in Epidemiological Studies. *Hum. Mol. Genet.* 23 (R1), R89–R98. doi:10.1093/hmg/ddu328

de Alcantara, C. C., Reiche, E. M. V., and Simão, A. N. C. (2021). Cytokines in Psoriasis. *Adv. Clin. Chem.* 100, 171–204. doi:10.1016/bs.acc.2020.04.004

de Groot, M., Teunissen, M. B. M., Ortonne, J. P., Lambert, J. R., Naeyaert, J. M., Picavet, D. I., et al. (2007). Expression of the Chemokine Receptor CCR5 in Psoriasis and Results of a Randomized Placebo Controlled Trial with a CCR5 Inhibitor. *Arch. Dermatol Res.* 299 (7), 305–313. doi:10.1007/s00403-007-0764-7

Deng, Y., Chang, C., and Lu, Q. (2016). The Inflammatory Response in Psoriasis: a Comprehensive Review. *Clin. Rev. Allerg. Immunol.* 50 (3), 377–389. doi:10.1007/s12016-016-8535-x

Dong, Y., Hu, H., Fu, D., Zheng, S., Wang, Q., K C, K., et al. (2021). Serum Expression of IL-33 and ST2 in Patients with Psoriasis Vulgaris. *Arch. Iran. Med.* 24 (9), 689–695. doi:10.34172/aim.2021.99

Duarte, G. V., Boeira, V., Correia, T., Porto-Silva, L., Cardoso, T., Macedo, M. N., et al. (2015). Osteopontin, CCL5 and CXCL9 Are Independently Associated with Psoriasis, Regardless of the Presence of Obesity. *Cytokine* 74 (2), 287–292. doi:10.1016/j.cyto.2015.04.015

Egger, M., Smith, G. D., and Phillips, A. N. (1997). Meta-analysis: Principles and Procedures. *BMJ* 315 (7121), 1533–1537. doi:10.1136/bmj.315.7121.1533

Emdin, C. A., Khera, A. V., and Kathiresan, S. (2017). Mendelian Randomization. *JAMA* 318 (19), 1925–1926. doi:10.1001/jama.2017.17219

Gilliet, M., Conrad, C., Geiges, M., Cozzio, A., Thürlimann, W., Burg, G., et al. (2004). Psoriasis Triggered by Toll-like Receptor 7 Agonist Imiquimod in the Presence of Dermal Plasmacytoid Dendritic Cell Precursors. *Arch. Dermatol* 140 (12), 1490–1495. doi:10.1001/archderm.140.12.1490

Goebeler, M., Toksoy, A., Spandau, U., Engelhardt, E., Bröcker, E.-B., and Gillitzer, R. (1998). The C-X-C Chemokine Mig Is Highly Expressed in the Papillae of Psoriatic Lesions. *J. Pathol.* 184 (1), 89–95. doi:10.1002/(sici)1096-9896(199801)184:1<89::aid-path975>3.0.co;2-z

Hammerberg, C., Arend, W. P., Fisher, G. J., Chan, L. S., Berger, A. E., Haskill, J. S., et al. (1992). Interleukin-1 Receptor Antagonist in Normal and Psoriatic Epidermis. *J. Clin. Invest.* 90 (2), 571–583. doi:10.1172/JCI115896

Harper, E. G., Guo, C., Rizzo, H., Lillis, J. V., Kurtz, S. E., Skorcheva, I., et al. (2009). Th17 Cytokines Stimulate CCL20 Expression in Keratinocytes *In Vitro* and *In Vivo*: Implications for Psoriasis Pathogenesis. *J. Investigative Dermatology* 129 (9), 2175–2183. doi:10.1038/jid.2009.65

Hawkes, J. E., Chan, T. C., and Krueger, J. G. (2017). Psoriasis Pathogenesis and the Development of Novel Targeted Immune Therapies. *J. Allergy Clin. Immunol.* 140 (3), 645–653. doi:10.1016/j.jaci.2017.07.004

Hemani, G., Bowden, J., and Davey Smith, G. (2018). Evaluating the Potential Role of Pleiotropy in Mendelian Randomization Studies. *Hum. Mol. Genet.* 27 (R2), R195–R208. doi:10.1093/hmg/ddy163

Johansen, C., Rittig, A. H., Mose, M., Bertelsen, T., Weimar, I., Nielsen, J., et al. (2017). STAT2 Is Involved in the Pathogenesis of Psoriasis by Promoting CXCL11 and CCL5 Production by Keratinocytes. *PLoS One* 12 (5), e0176994. doi:10.1371/journal.pone.0176994

Johansen, C., Usher, P. A., Kjellerup, R. B., Lundsgaard, D., Iversen, L., and Kragballe, K. (2009). Characterization of the Interleukin-17 Isoforms and Receptors in Lesional Psoriatic Skin. *Br. J. Dermatol* 160 (2), 319–324. doi:10.1111/j.1365-2133.2008.08902.x

Joshi, N., Narang, T., Dogra, S., and Chhabra, S.(2019). Circulating Levels of Chemokines in Patients with Psoriasis Vulgaris and Their Association with Disease Severity: A Case-Control Study from North India. *Indian J. Dermatol Venereol. Leprol.* 85 (5), 567. doi:10.4103/ijdvl.IJDVL_392_18

Kanda, N., Koike, S., and Watanabe, S. (2005). IL-17 Suppresses TNF-α-Induced CCL27 Production through Induction of COX-2 in Human Keratinocytes. *J. Allergy Clin. Immunol.* 116 (5), 1144–1150. doi:10.1016/j.jaci.2005.08.014

Kaufman, B. P., and Alexis, A. F. (2018). 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* 19 (3), 405–423. doi:10.1007/s40257-017-0332-7

Kim, H. J., Kim, S. H., Park, J., Lee, M., Kim, D. S., and Lee, M.-G. (2016). Up-regulation of Receptor Antagonist Interleukin-1 Family Members in Psoriasis and Their Regulation by Pro-inflammatory Cytokines. *J. Dermatological Sci.* 82 (3), 204–206. doi:10.1016/j.jdermsci.2016.02.003

Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N., and Davey Smith, G. (2008). Mendelian Randomization: Using Genes as Instruments for Making Causal Inferences in Epidemiology. *Stat. Med.* 27 (8), 1133–1163. doi:10.1002/sim.3034

Levy, J. A. (2009). The Unexpected Pleiotropic Activities of RANTES. *J. Immunol.* 182 (7), 3945–3946. doi:10.4049/jimmunol.0990015

Machiela, M. J., and Chanock, S. J. (2015). LDlink: a Web-Based Application for Exploring Population-specific Haplotype Structure and Linking Correlated Alleles of Possible Functional Variants. *Bioinformatics* 31 (21), 3555–3557. doi:10.1093/bioinformatics/btv402

Marques, R. E., Guabiraba, R., Russo, R. C., and Teixeira, M. M. (2013). Targeting CCL5 in Inflammation. *Expert Opin. Ther. Targets* 17 (12), 1439–1460. doi:10.1517/14728222.2013.837886

Nedoszytko, B., Sokołowska-Wojdyło, M., Ruckemann-Dziurdzińska, K., Roszkiewicz, J., and Nowicki, R. J. (2014). Chemokines and Cytokines Network in the Pathogenesis of the Inflammatory Skin Diseases: Atopic Dermatitis, Psoriasis and Skin Mastocytosis. *Postepy Dermatol Alergol.* 2 (2), 84–91. doi:10.5114/pdia.2014.40920

Nestle, F. O., Conrad, C., Tun-Kyi, A., Homey, B., Gombert, M., Boyman, O., et al. (2005). Plasmacytoid Predendritic Cells Initiate Psoriasis through Interferon-α Production. *J. Exp. Med.* 202 (1), 135–143. doi:10.1084/jem.20050500

Pierce, B. L., and Burgess, S. (2013). Efficient Design for Mendelian Randomization Studies: Subsample and 2-sample Instrumental Variable Estimators. *Am. J. Epidemiol.* 178 (7), 1177–1184. doi:10.1093/aje/kwt084

Ramirez-Carrozzi, V., Sambandam, A., Luis, E., Lin, Z., Jeet, S., Lesch, J., et al. (2011). IL-17C Regulates the Innate Immune Function of Epithelial Cells in an Autocrine Manner. *Nat. Immunol.* 12 (12), 1159–1166. doi:10.1038/ni.2156

Rateb, A. A., Fawzi, M. M. T., Abdel Hay, R. M., Mohammed, F. N., and Amr, K. S. (2012). Reduction of RANTES Expression in Lesional Psoriatic Skin after Narrow Band Ultraviolet Therapy: a Possible Marker of Therapeutic Efficacy. *Eur. J. Dermatol* 22 (4), 481–487. doi:10.1684/ejd.2012.1730

Rendon, A., and Schäkel, K. (2019). Psoriasis Pathogenesis and Treatment. *Int. J. Mol. Sci.* 20 (6), 1475. doi:10.3390/ijms20061475

Smith, G. D., and Ebrahim, S. (2004). Mendelian Randomization: Prospects, Potentials, and Limitations. *Int. J. Epidemiol.* 33 (1), 30–42. doi:10.1093/ije/dyh132

Sobota, R. S., Shriner, D., Kodaman, N., Goodloe, R., Zheng, W., Gao, Y.-T., et al. (2015). Addressing Population-specific Multiple Testing Burdens in Genetic Association Studies. *Ann. Hum. Genet.* 79 (2), 136–147. doi:10.1111/ahg.12095

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., et al. (2015). UK Biobank: an Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med.* 12 (3), e1001779. doi:10.1371/journal.pmed.1001779

Verbanck, M., Chen, C.-Y., Neale, B., and Do, R. (2018). Detection of Widespread Horizontal Pleiotropy in Causal Relationships Inferred from Mendelian Randomization between Complex Traits and Diseases. *Nat. Genet.* 50 (5), 693–698. doi:10.1038/s41588-018-0099-7

von Csiky-Sessoms, S., and Lebwohl, M. (2019). What's New in Psoriasis. *Dermatol. Clin.* 37 (2), 129–136. doi:10.1016/j.det.2018.11.001

Zablotna, M., Sobjanek, M., Purzycka-Bohdan, D., Szczerkowska-Dobosz, A., Nedoszytko, B., and Nowicki, R. (2016). The -2518 A/GMCP-1and -403 G/ARANTESpromoter Gene Polymorphisms Are Associated with Psoriasis Vulgaris. *Clin. Exp. Dermatol* 41 (8), 878–883. doi:10.1111/ced.12937

Zhou, W., Nielsen, J. B., Fritsche, L. G., Dey, R., Gabrielsen, M. E., Wolford, B. N., et al. (2018). Efficiently Controlling for Case-Control Imbalance and Sample Relatedness in Large-Scale Genetic Association Studies. *Nat. Genet.* 50 (9), 1335–1341. doi:10.1038/s41588-018-0184-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.

Copyright © 2022 Zhao, Zhang, Liu, Tang, Wang, Wang, Wu, Yang, Li and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 16 June 2022
doi: 10.3389/fgene.2022.921808

Clinical Characteristics, *In Silico* Analysis, and Intervention of Neonatal-Onset Inflammatory Bowel Disease With Combined Immunodeficiency Caused by Novel *TTC7A* Variants

Yun-e Chen[1†], Jingfang Chen[2†], Wenxing Guo[1], Yanhong Zhang[3], Jialing Li[2], Hui Xie[1], Tong Shen[1], Yunsheng Ge[4], Yanru Huang[4], Wenying Zheng[5] and Mei Lu[1]**

[1]Department of Pediatrics, Women and Children's Hospital, School of Medicine, Xiamen University, Xiamen, China, [2]Department of Gastroenterology, Xiamen Branch of the Children's Hospital of Fudan University (Xiamen Children's Hospital), Xiamen, China, [3]Department of Ultrasound Medicine, Women and Children's Hospital, School of Medicine, Xiamen University, Xiamen, China, [4]Prenatal Diagnostic Center Laboratory, Women and Children's Hospital, School of Medicine, Xiamen University, Xiamen, China, [5]Genokon Institute of Medical Science and Laboratory, Xiamen, China

We aimed to explore the genotypic and phenotypic characteristics of neonatal-onset inflammatory bowel disease (IBD) with combined immunodeficiency due to *TTC7A* mutation. We examined the clinical manifestations, imaging results, endoscopic and histological findings, interventions, and prognosis of a proband with neonatal-onset IBD and performed biochemical analyses, whole-exome sequencing (WES), and *in silico* analysis. Our proband developed severe early-onset diarrhea, malnutrition, electrolyte imbalance, dehydration, and recurrent infections after birth. Radiographic and ultrasonic images showed no specific manifestations. Endoscopic and histological examination revealed chronic inflammation. Immune function examination indicated immunodeficiency. WES identified compound heterozygous *TTC7A* mutations (c.2355+4A>G, c.643G>T) in the proband. In the expression analysis, no abnormal splicing in the *TTC7A* sequence was observed due to the c.2355+4A>G mutation; however, the mRNA expression was reduced. The proband's condition did not improve after treatment with methylprednisolone or leflunomide. The proband died when treatment was stopped at the age of 5 months and 19 days. Compound heterozygous mutations (c.2355+4A>G, c.643G>T) in the *TTC7A* gene are described and verified for the first time. Our report expands the phenotypic spectrum of *TTC7A* mutations and the genotypic spectrum of very early-onset IBD with combined immunodeficiency.

Keywords: inflammatory bowel disease, immunodeficiency, neonatal, *TTC7A*, intervention

OPEN ACCESS

Edited by:
Yunqing Ren,
Zhejiang University, China

Reviewed by:
Lidan Hu,
Zhejiang University, China
Xinyi Meng,
Tianjin Medical University, China

****Correspondence:***
Yanru Huang
huangyanruxm@163.com
Mei Lu
lm800529@126.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: *16 April 2022*
Accepted: *30 May 2022*
Published: *16 June 2022*

Citation:
Chen Y-e, Chen J, Guo W, Zhang Y, Li J, Xie H, Shen T, Ge Y, Huang Y, Zheng W and Lu M (2022) Clinical Characteristics, In Silico Analysis, and Intervention of Neonatal-Onset Inflammatory Bowel Disease With Combined Immunodeficiency Caused by Novel TTC7A Variants. Front. Genet. 13:921808. doi: 10.3389/fgene.2022.921808

INTRODUCTION

Very early onset inflammatory bowel disease (VEOIBD) encompasses a group of diseases in children aged less than 6 years (Kelsen et al., 2019); it has a unique clinical presentation with severe colonic disease in infancy (Griffiths, 2004; Muise et al., 2012a). Mutations in *IL10RA/B*, *IL106*, *XIAP7*, *ADAM178*, *NCF49*, *NCF2/RAC210*, and *TTC7A* cause a severe form of VEOIBD, with symptoms developing in infancy (Glocker et al., 2009; Matute et al., 2009; Glocker et al., 2010; Blaydon et al., 2011; Worthey et al., 2011; Muise et al., 2012b; Kelsen and Sullivan, 2017). *TTC7A*, a member of the tetratricopeptide repeat (TPR) domain-containing proteins, has multiple functions in cell cycle control, phosphate turnover, and protein transport and secretion (White et al., 2005). *TTC7A* gene mutations occur in several patients with multiple intestinal atresia (MIA) and gastrointestinal defects, immunodeficiency syndrome, and inflammatory bowel disease (IBD) with or without immunodeficiency; they are inherited in an autosomal recessive manner (OMIM: 609332). Approximately 50 cases of TTC7A deficiency have been reported worldwide, most of which are MIA with combined immunodeficiency (CID), and only a few are IBD cases (Jardine et al., 2019a). Missense *TTC7A* mutations often result in clinical features of VEOIBD, and truncating *TTC7A* mutations (nonsense, frameshifts, or large deletions) are associated with greater morbidity and mortality in MIA-CID patients (Samuels et al., 2013; Avitzur et al., 2014; Bigorgne et al., 2014). Patients with mutations affecting the splice acceptor sites present with intestinal atresia, increased intestinal cell apoptosis, and severe bowel inflammation (Samuels et al., 2013); however, it is difficult to establish a significant correlation between the genotype and phenotype. Most patients were diagnosed within 1 year after birth; only two patients were diagnosed in the neonatal period (Jardine et al., 2019a). Neonatal-onset IBD often exhibits a poor response to conventional treatments in contrast to adult-onset IBD (Heyman et al., 2005).

In this study, we describe the clinical presentation, imaging results, pathological features, functional analysis, treatment, and prognosis of a proband with neonatal-onset IBD caused by two novel *TTC7A* compound heterozygous mutations. There was no improvement in the proband's condition after treatment with methylprednisolone or leflunomide. This is the first known case of *TTC7A* mutations in mainland China and it extends the pathogenic spectrum of this gene.

MATERIALS AND METHODS

Proband

We retrospectively studied a proband from the Women and Children's Hospital, School of Medicine, Xiamen University, China. All relevant clinical information was collected, including clinical progression, laboratory findings, imaging, and endoscopic and histological characteristics. Stool electrolytes were detected using biochemical testing methods.

TABLE 1 | Lymphocyte subsets and immunoglobulin levels of the proband.

	Day 52[a]	Day 65	Day 78	Day 105[b]	References
IgA (g/L)	0.35	<0.28	<0.28	0.06	0.05–0.60
IgG (g/L)	5.51	3.93	1.4	5.1	2.75–7.50
IgM (g/L)	0.26	0.53	0.34	0.29	0.10–0.70
C3 (g/L)	0.65	0.98	1.04	0.55	0.70-1.40
C4 (g/L)	0.15	0.15	0.18	0.14	0.10–0.40
$CD3^+T$ (10^9/L)	0.77	—	—	2.04	1.85–4.02
$CD3^+CD4^+T$ (10^9/L)	0.56	—	—	0.32	1.33–3.11
$CD3^+CD8^+T$ (10^9/L)	0.21	—	—	1.45	0.66–1.15
NK (10^9/L)	0.07	—	—	5.16	0.27–0.73
B (10^9/L)	0.50	—	—	16.93	0.34–2.09

[a]*2 days after 5 g IgG infusion.*
[b]*3 days after 5 g IgG infusion.*

Whole-Exome Sequencing and Variant Classification

Peripheral blood samples of the proband, his parents, and control (a normal peer) were collected, and trio-based WES was performed at Guangzhou KingMed Diagnostics Group Co. Ltd. (Guangzhou, China). Genomic DNA was extracted using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany), and mechanical shearing was performed using Covaris AFA-energetics® (Covaris Inc., Woburn, MA, United States) according to the manufacturer's instructions. DNA library preparation and adaptor hybridization were performed according to the standard procedures of the Agencourt AMPure XP Beads (Beckman Coulter Inc., Brea, CA, United States). The DNA library was used to capture and collect DNA from the target exons and adjacent splice sites, using the Agilent SureSelect Human All Exon V6 kit (58 M). Finally, PE150 + 150 sequencing was performed using Illumina NovaSeq 6000 platform (Illumina, Inc., San Diego, CA, United States). The reads were mapped to the reference human genome (UCSC hg19).

Primers were designed to amplify the candidate regions of the mutations identified using WES. Subsequently, Sanger sequencing using ABI PRISM 3130XL gene analyzer (Applied Biosystems, California, United States) was performed.

In Silico Analysis

All variants were compared with the data in the 1000 Genomes Project (https://www.internationalgenome.org), Exome Aggregation Consortium (ExAC) (https://exac.broadinstitute.org/), and dbSNP (https://www.ncbi.nlm.nih.gov/snp/) databases. Variants with greater than 0.01 minor allele frequency in the control databases were excluded. Pathogenicity predictions for proteins were performed using Revel, SIFT, PolyPhen2, PROVEAN_pred, LRT_pred, Mutation Taster, etc.

Analysis of *TTC7A* Expression

Total RNA was extracted from the peripheral blood lymphocyte cell lines using PowerUpTM SYBR Green Master Mix Kit (Life Technologies). The mRNA was reverse transcribed to cDNA using RevertAid First Strand cDNA Synthesis Kit (Thermo

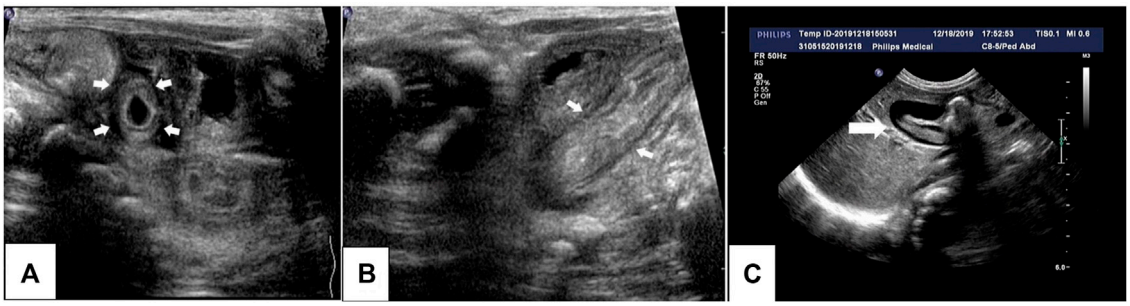


FIGURE 1 | Ultrasonography of the proband's abdomen. Thickening of the small intestine and colon, enhanced echogenicity of the mucous membrane and submucosa, and effusion in the colon are visible **(A,B)**. Low-echogenicity area in the gallbladder (bile sludge) **(C)** can be seen, which disappeared the next day.

Scientific). The region near the splice site was amplified using the primers, 5′-TCCCCACTTCTCACTCAGTACTC-3′ and 5′-GTG GCACGTACTCTGCCTCT-3′ (product length: 218 bp), followed by agarose gel electrophoresis and sequencing through Sanger sequencing. Real-time quantitative PCR (qPCR) was performed to verify whether the splice site mutation affected the mRNA expression. The relative standard curve method was used to analyze the expression level. The expression was normalized to that of*ACTB* in the same sample; three biological repeats were measured. The real-time PCR primers used were: *TTC7A*, 5′-ATGCATAGCCTGGGT CTGAT-3′ and 5′-GTGGCACGTACTCTGCCTCT-3′ (product length: 96 bp); *ACTB*, 5′-TGGTGCCAGATTTTC TCCA-3′ and 5′-GGCATGGGTCAGAAGGATT-3′ (product length: 128 bp).

Informed Consent

Written informed consent for genetic testing and the publication of research data was obtained from the proband's parents. The protocol was approved by the Women and Children's Hospital Ethics Committee (KY-2020-011).

RESULTS

Clinical Features of the Proband

The proband, a male infant conceived *via in vitro* fertilization, was born to non-consanguineous Chinese parents at 35 weeks and 6 days of gestation. The clinical characteristics of his parents were normal; and the mother and father were aged 27 and 29 years, respectively. No complications were observed during the prenatal period. No exposure to toxic or harmful substances during pregnancy was reported. There was no family history of genetic diseases. His birth weight and length were 2,800 g and 50 cm, respectively. He was breastfed after birth and presented with diarrhea with yellow, watery, and bloody stool 5–10 times per day. He was admitted to a tertiary children's hospital; however, he showed no improvement after 1 week of antimicrobial treatment. He was admitted to our hospital for further treatment, at the age of 1 month and 8 days, with a weight of 2,600 g, head circumference of 33 cm, and body length of 50 cm, without dysmorphic features. The proband had malnutrition, moderate dehydration, and severe perianal dermatitis.

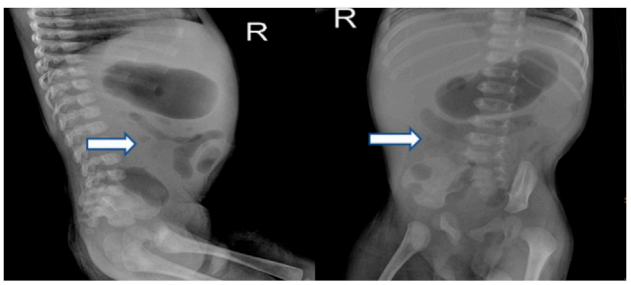


FIGURE 2 | X-ray of the proband's abdomen revealed stiffness in the intestines and low volume of intestinal gas.

Because of the large quantity of watery stool (approximately 300–940 ml/day), total parenteral nutrition, amino acid milk powder, deep hydrolyzed protein milk powder, or breast milk were administered after admission; however, there was no improvement in his condition. The stool was considered to be high-output secretory diarrhea. Severe hypokalemia, hyponatremia, and metabolic acidosis occurred, and the proband recovered after total parenteral nutrition every day. The erythrocyte sedimentation rate was normal; however, the ferritin levels were >1,500 ng/ml (reference value <200 ng/ml). Interleukin (IL)-6 levels reached 74.6 pg/ml (reference value ≤5.9 pg/ml), and the tumor necrosis factor-α level was 39.2 pg/ml (reference value ≤ 8.1 pg/ml). A diagnosis of IBD was suspected.

During hospitalization, the proband had recurrent fever; blood immune function investigation indicated CID (**Table 1**). The test for cytomegalovirus antibody in the blood was negative; however, the amounts of cytomegalovirus DNA in the blood and urine were 4.12×10^6 copies/ml and 1.47×10^6 copies/ml, respectively. Cytomegalovirus and *Pneumocystis carinii* were detected in the bronchoalveolar lavage fluid through metagenomic detection; *Candida albicans* was detected in urine culture and sputum culture. The risk factors for these infections include malnutrition, CID, long hospital stays, and long-term use of antibiotics and hormones.

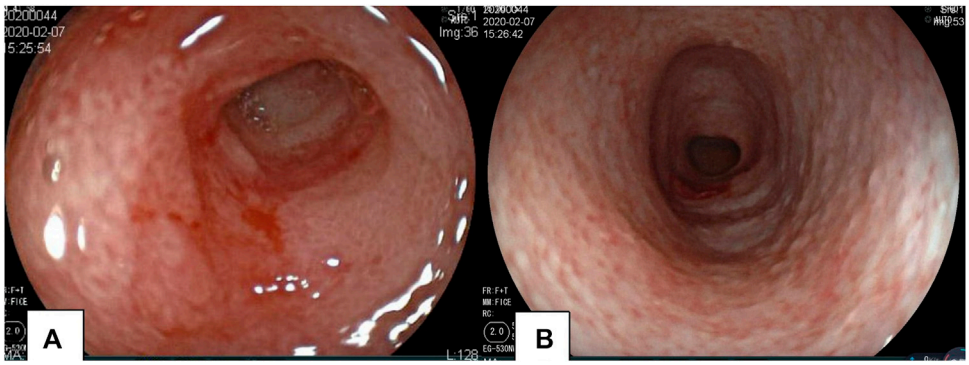


FIGURE 3 | Gastroscopy and colonoscopy of the proband. Mucosal erosion and edema of the pylorus revealed by gastroscopy **(A)**. Colonoscopy revealed the mucosa of the whole colon to be rough; the surface was scattered with a white pseudomembrane. Part of the intestinal mucosa was eroded, and the intestine was slightly narrow **(B)**.

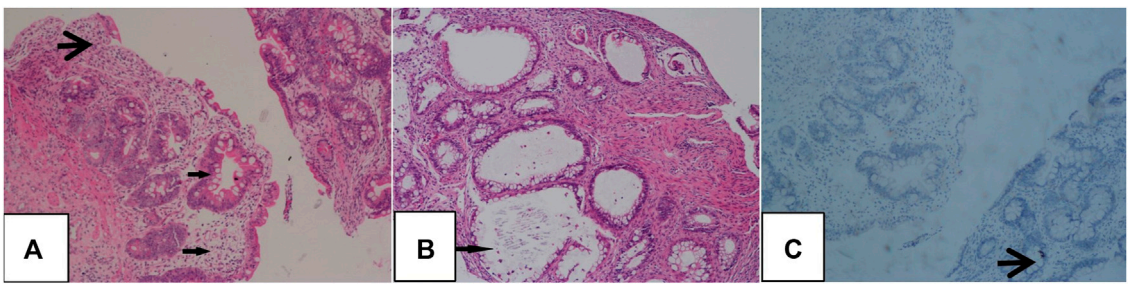


FIGURE 4 | Histopathological findings of the proband's intestine. Chronic mild superficial inflammatory changes are seen in the esophagus and stomach. The duodenal crypt was atrophied, and plasma cells were scattered in the interstitium **(A)**. Twisted colon crypts, mucus hypersecretion **(B)**, focal epithelial hyperplasia, and basal plasmacytosis are evident in the large intestinal mucosa **(C)**.

The proband received cefotaxime, piperacillin-tazobactam, cefoperazone-sulbactam, meropenem, ganciclovir, co-trimoxazole, fluconazole, voriconazole, and intravenous immunoglobulin (IVIG). Methylprednisolone 1–2 mg/kg/day was given intravenously for 11 days. He received leflunomide (the dose increased over time from 0.5 to 1.0 mg/days) for 4 weeks, commencing at the age of 4 months 23 days. Although his infections were under control after comprehensive treatment, he still presented with refractory diarrhea, and his weight remained in the range of 2,300–3,400 g after 4 months. The infant died when his parents discontinued treatment at the age of 5 months and 19 days.

Imaging Features and Endoscopic and Histological Characteristics

The proband underwent abdominal ultrasound, radiography, gastroscopy, colonoscopy, and biopsy pathological analyses (**Figures 1–4**). Gastroscopy and colonoscopy examination were performed on days 98 and 100, respectively.

WES and Variant Classification

Trio-based WES identified two novel heterozygous variants in *TTC7A*: a mutation in exon 4, c.643G>T, inherited from his father, and another in exon 19, c.2355+4A>G inherited from his mother. According to the American College of Medical Genetics and Genomics and the Association for Molecular Pathology variant-interpretation guidelines (Richards et al., 2015) and the work groups from ClinGen (https://clinicalgenome.org/), the nonsense mutation c. 643G>T causes premature termination of the protein at amino acid 215 (p. Glu215X); the variant is classified as a pathogenic variant (PVS1 + PM2-supporting + PP3). The intronic allele c.2355+4A>G is predicted to affect mRNA splicing of TTC7A transcripts, according to dbscSNV_ADA and dbscSNV_RF; the variant is classified as a variable of uncertain significance (PM2-supporting + PP3; **Figure 5**).

TTC7A Expression Analysis

No abnormal splicing bands were observed in the agarose gel electrophoresis; no abnormal splicing sequences were detected using Sanger sequencing. However, the expression of *TTC7A* mRNA decreased in the proband ($p < 0.01$), which could have resulted in a decrease in the protein levels. However, no aberrant expression of TTC7A was detected in the parents (**Figures 6–8**). The decrease in mRNA expression could lead to decreased protein expression or the production of the wrong protein (Chen et al., 2013; Bigorgne et al., 2014).

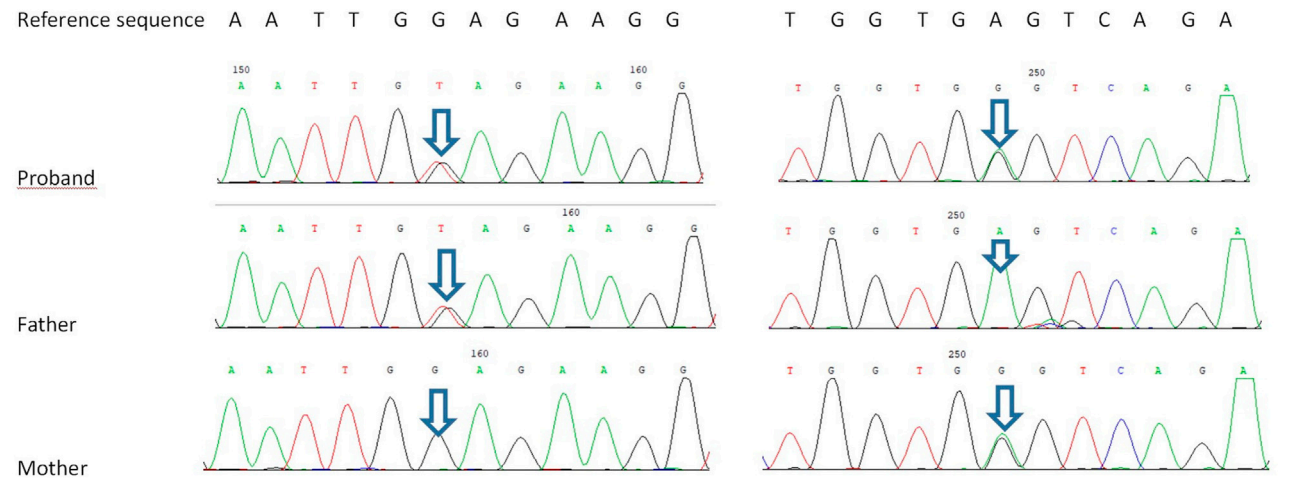


FIGURE 5 | Sanger sequencing validation of the *TTC7A* variants.

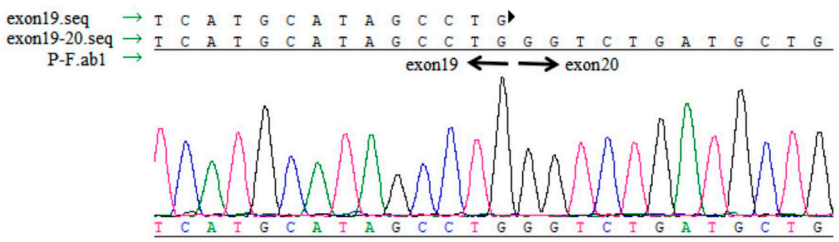


FIGURE 6 | Complementary DNA sequencing of the proband.

DISCUSSION

The global incidence rate of IBD is 37.5/100,000 annually. In the United States, the incidence of IBD was the highest in adults (18–59 years) and older adults (>60 years) from 2005 to 2015 and the lowest in pediatric patients (0–17 years) (Keyashian et al., 2019). In a study of 1,412 pediatric cases, VEOIBD accounted for approximately 3% of cases, and infantile-onset IBD accounted for only 1%. The incidence of VEOIBD remained stable over the study period (1988–2011); however, the incidence of EOIBD increased (Bequet et al., 2017). These results suggest a probable genetic origin for VEOIBD, whereas the increased incidence in older children may be linked to environmental factors (Bequet et al., 2017). The most important clinical sign of monogenic IBD is the young age at onset. In this study, the case was initially thought to be related to genetic factors because of the very early onset and severe symptoms.

Neonatal or infantile-onset IBD is often associated with primary immune deficiency disorders, especially those caused by mutations in *IL-10*, *XIAP*, *NCF2*, *IPEX*, and *TTC7A* (Kelsen and Sullivan, 2017). *TTC7A* has 20 exons and nine TPR domains, and mutations in exons 2, 7, and 20 are relatively common (Lien et al., 2017). TPR domains are structurally conserved helix-turn-helix motifs that play an important role in protein scaffolding and have large surface areas that accommodate multiple protein interactions (Blatch and Lässle, 1999; Zeytuni and Zarivach, 2012). No case of *TTC7A* mutation was reported in mainland China, before this case. Our proband had two novel *TTC7A* mutations, c.643G>T in exon 4, inherited from his father, and c.2355+4A>G in exon 19, inherited from his mother, which was not described in the Human Gene Mutation Database. *TTC7A* mutation exhibits an autosomal recessive inheritance pattern; the proband's parents have normal phenotypes. The c.643G>T mutation, a nonsense mutation in exon 9, results in premature termination of the protein and is predicted to cause loss of function. qPCR showed that the c.2355+4A>G mutation decreased *TTC7A* mRNA expression. We believed that this intron mutation would affect splicing; however, the experimental results revealed that this was not the case.

To date, few studies have performed functional analysis of the specific variants of *TTC7A*. We did not perform immunostaining analysis for assessing the proteins; however, we speculate that the c.2355+4A>G mutation could affect the TPR domain, which binds and recruits phosphatidylinositol 4 (PI4)-kinase III alpha to the plasma membrane. This process facilitates the synthesis of mutant PI4-phosphate affecting the TPR domain, resulting in negative phenotypes and poor outcomes (Lien et al., 2017). Studies on truncating *TTC7A* mutations suggested that nonsense-mediated decay of *TTC7A* mRNA transcripts showed an obvious loss of protein, consistent with the increased phenotypic severity in

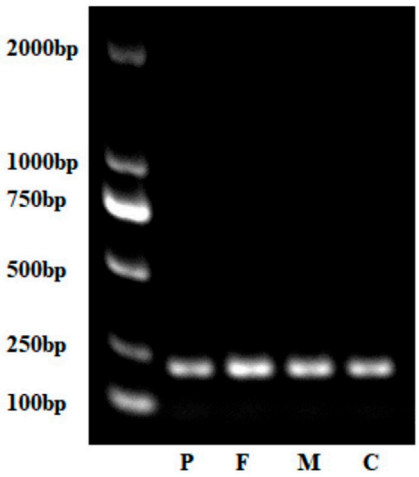


FIGURE 7 | Gel electrophoresis of complementary DNA of the proband, parents, and control (normal peer). P, proband; F, father; M, mother; C, control.

patients with MIA--CID, whereas non-truncating mutations tended to present as VEOIBD (Samuels et al., 2013; Avitzur et al., 2014). Nonsense mutations cause severe intestinal diseases (e.g., MIA) and CID; however, missense mutations mainly present as IBD and mild intestinal disease (Lawless et al., 2017; Neves et al., 2018). Our proband had a truncated mutation that affected the TPR domain, which manifested as neonatal-onset IBD and CID, but not as MIA. It is difficult to attribute complex phenotypes to physical alterations in TTC7A.

Fewer than 10% of patients with *TTC7A* mutations demonstrate VEOIBD with secretory diarrhea, chronic intestinal inflammation, lymphocytopenia, and/or hypoglobulinemia (Lien et al., 2017; Jardine et al., 2019a). Patients with VEOIBD are more likely to present with ileal involvement and isolated colonic involvement. Rectal bleeding and mucous stools with weight loss and underlying primary immunodeficiency appear to be more common in VEOIBD cases with monogenic causes (Bequet et al., 2017; Kelsen et al., 2019; Shim, 2019).

Radiography and ultrasonography are the most commonly used methods of examination in pediatric patients, especially before the diagnosis of IBD. No specificity or significant differences between pediatric and adult IBD patients were identified (Horsthuis et al., 2008). Imaging studies of VEOIBD are relatively rare. Hepatosplenomegaly, significant intestinal wall thickening, and mildly dilated bowels are reported in neonatal IBD (Barber et al., 2018). In our proband, abdominal radiography revealed stiffness of the intestines and decreased intestinal gas volume. Gastrointestinal ultrasonography revealed thickening of the bowel wall, reduction in intestinal gas volume, and fluid accumulation in the colon. All these imaging findings were consistent with the pathologic changes of apoptosis, twisted crypts, and hypersecretion of mucus in the colon. It is very difficult to perform gastroenteroscopy (GE) in the early neonatal or infant stages because of the severe risk conditions, poor bowel preparation, and technical failure. The proband underwent GE at approximately 3 months of age, and chronic mild superficial inflammation in the esophagus, stomach, and duodenum, twisted crypts, and hypersecretion of mucus in the colon were observed. These features, unlike that in the atypical tissues observed in older children and adult ulcerative colitis or Crohn's disease, mainly involve the colon. Compared to that in adult IBD patients, extensive ileocolonic inflammation is more predominant in pediatric IBD patients (Peloquin et al., 2016). The most common endoscopic findings are mucosal bleeding among patients with VEOIBD and visual ileitis or ileocolonic bleeding in older patients with IBD (Conrad et al., 2019).

Severe chronic structural changes, increased frequency of apoptosis, blunt villi in the small intestine, and eosinophils in the surface epithelium, crypts, and lamina propria reported in VEOIBD are discovered in some primary immunodeficiencies (Conrad et al., 2019). T-cell maturation disruption and lymphocytopenia owing to a hypoplastic thymus and low blood immunoglobulin levels caused by increases in the relative levels of transitional B cells and B-cell receptor assembly and organelle

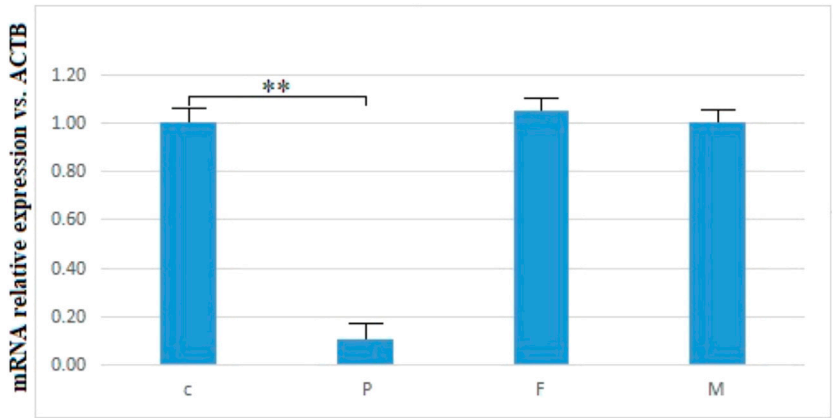


FIGURE 8 | mRNA relative expression of *TTC7A* vs. *ACTB* of the proband, parents, and control. $^{**}p < 0.01$. P, proband; F, father; M, mother; C, control.

synthesis in activated B cells are speculated (Avitzur et al., 2014; Woutsas et al., 2015). Therefore, immunodeficiency is closely related to VEOIBD; immune function should be tested as soon as possible during diagnosis. According to the data from our case, the phenotype involves extensive lesions in the stomach, duodenum, small intestine, and colon, leading to very early onset, severe symptoms, and immunodeficiency.

Most monogenic IBD patients are resistant to conventional medical treatment (Heyman et al., 2005; Kelsen et al., 2019; Shim, 2019), as are patients with neonatal-onset IBD with *TTC7A* deficiency. Currently, treatment options include early intestinal resection (in MIA), parenteral nutrition, and regular immunoglobulin infusion (Kammermeier et al., 2016). Total or partial parenteral nutrition has been reported in many patients with *TTC7A* mutations presenting with MIA, gastrointestinal defects, and IBD (Avitzur et al., 2014; Bigorgne et al., 2014; Lemoine et al., 2014); early application of parenteral nutrition is necessary because of intestinal malabsorption and excessive intestinal loss. In our study, parenteral nutritional support was initiated early.

VEOIBD shows poor response to immunosuppressives, steroids, and biologics (Jardine et al., 2019a). Hematopoietic stem cell transplantation treatment has been reported for monogenic IBD, which may restore immunity and increase survival in immunodeficient patients; however, it does not appear to improve the phenotypes associated with bowel epithelial defects (Jardine et al., 2019a). The survival rate of patients with *TTC7A* deficiency treated with HSCT did not improve (Kammermeier et al., 2016; Fayard et al., 2018). Intestinal and liver transplantation restores intestinal and immune functions in a child with MIA-CID (Gilroy et al., 2004). Unfortunately, intestinal transplantation is not feasible because of the limitations of donor factors. *In vitro* studies show that Rho A inhibitors (Y-27632) and leflunomide are effective treatments (Notarangelo, 2014; Jardine et al., 2019b); however, no *TTC7A*-deficient patients have received leflunomide. After we obtained consent from the parents, we started leflunomide treatment for 4 weeks but without any success. The efficacy of leflunomide *in vivo* is uncertain. Leflunomide is mainly used for treating rheumatoid arthritis. Leflunomide was effective in an *in vivo* zebrafish study on TTC7A deficiency; however, this drug was not effective in our proband. Further elucidation of the pathogenesis and therapeutic targets in TTC7A-deficiency through the accumulation of clinical experience is warranted. The infant died when treatment was stopped at the age of 5 months and 19 days.

The *TTC7A*-deficient genotype may be closely related to prognosis. The average survival of patients with *TTC7A* nonsense mutations is less than 12 months. Some patients with biallelic missense mutations that do not result in autoimmune disorders or involve TPR domains have a relatively better prognosis and can survive to adulthood (Lawless et al., 2017; Neves et al., 2018).

CONCLUSION

Two novel compound heterozygous mutations in *TTC7A* were identified for the first time in mainland China. These mutations manifested as neonatal-onset IBD-CID with a poor prognosis. No specific manifestations were found with imaging or endoscopy. Studying the genotype and clinical phenotype of TTC7A deficiency requires the accumulation of more clinical data and in-depth functional research.

DATA AVAILABILITY STATEMENT

The datasets for this article are not publicly available due to concerns regarding participant/patient anonymity. Requests to access the datasets should be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Women and Children's Hospital Ethics Committee, School of Medicine, Xiamen University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

ML and YH conceived and planned this study and take full responsibility for this study; Y-eC, JC, and ML drafted this manuscript; Y-eC, JC, ML, JL, WG, HX, and TS contributed to clinical analysis; ML, YG, YH, and WZ contributed to the interpretation of results and genetic data; YZ reviewed the images; and JC contributed to the analysis of the endoscopic and histological data. All authors have contributed and approved the final manuscript.

FUNDING

This study was supported by the Xiamen Science and Technology Guiding Project (3502Z20199078 to ML) and the backbone talent cultivation project of the Xiamen Children's Hospital in 2019 (CHP-2019-BT-B001 to JC).

ACKNOWLEDGMENTS

We are very grateful to the infant and his family for their cooperation and contribution to this study. We would like to acknowledge Ying Huang (Children's Hospital of Fudan University, Shanghai, China) for her guidance regarding treatment.

REFERENCES

Avitzur, Y., Guo, C., Mastropaolo, L. A., Bahrami, E., Chen, H., Zhao, Z., et al. (2014). Mutations in Tetratricopeptide Repeat Domain 7A Result in a Severe Form of Very Early Onset Inflammatory Bowel Disease. *Gastroenterology* 146, 1028–1039. doi:10.1053/j.gastro.2014.01.015

Barber, J. L., Shah, N., and Watson, T. A. (2018). Early Onset Inflammatory Bowel Disease - What the Radiologist Needs to Know. *Eur. J. Radiol.* 106, 173–182. doi:10.1016/j.ejrad.2018.07.001

Bequet, E., Sarter, H., Fumery, M., Vasseur, F., Armengol-Debeir, L., Pariente, B., et al. (2017). Incidence and Phenotype at Diagnosis of Very-Early-Onset Compared with Later-Onset Paediatric Inflammatory Bowel Disease: A Population-Based Study [1988-2011]. *J. Crohns Colitis* 11, 519–526. doi:10.1093/ecco-jcc/jjw194

Bigorgne, A. E., Farin, H. F., Lemoine, R., Mahlaoui, N., Lambert, N., Gil, M., et al. (2014). TTC7A Mutations Disrupt Intestinal Epithelial Apicobasal Polarity. *J. Clin. Invest.* 124, 328–337. doi:10.1172/jci71471

Blatch, G. L., and Lässle, M. (1999). The Tetratricopeptide Repeat: A Structural Motif Mediating Protein-Protein Interactions. *Bioessays* 21, 932–939. doi:10.1002/(sici)1521-1878(199911)21:11<932::aid-bies5>3.0.co;2-n

Blaydon, D. C., Biancheri, P., Di, W.-L., Plagnol, V., Cabral, R. M., Brooke, M. A., et al. (2011). Inflammatory Skin and Bowel Disease Linked to ADAM17 Deletion. *N. Engl. J. Med.* 365, 1502–1508. doi:10.1056/nejmoa1100721

Chen, R., Giliani, S., Lanzi, G., Mias, G. I., Lonardi, S., Dobbs, K., et al. (2013). Whole-Exome Sequencing Identifies Tetratricopeptide Repeat Domain 7A (TTC7A) Mutations for Combined Immunodeficiency with Intestinal Atresias. *J. Allergy Clin. Immunol.* 132, 656–664. doi:10.1016/j.jaci.2013.06.013

Conrad, M. A., Carreon, C. K., Dawany, N., Russo, P., and Kelsen, J. R. (2019). Distinct Histopathological Features at Diagnosis of Very Early Onset Inflammatory Bowel Disease. *Bowel Dis.* 13, 615–625. doi:10.1093/ecco-jcc/jjy212

Fayard, J., Collardeau, S., Bertrand, Y., Cordier, M. P., Malcus, C., Dubois, R., et al. (2018). TTC7A Mutation Must Be Considered in Patients with Repeated Intestinal Atresia Associated with Early Inflammatory Bowel Disease: Two New Case Reports and a Literature Review. *Arch. Pediatr.* S0929-693X (18), 30112. doi:10.1016/j.arcped.2018.05.006

Gilroy, R. K., Coccia, P. F., Talmadge, J. E., Hatcher, L. I., Pirruccello, S. J., Shaw, B. W., Jr., et al. (2004). Donor Immune Reconstitution after Liver-Small Bowel Transplantation for Multiple Intestinal Atresia with Immunodeficiency. *Blood* 103, 1171–1174. doi:10.1182/blood-2003-04-1187

Glocker, E.-O., Frede, N., Perro, M., Sebire, N., Elawad, M., Shah, N., et al. (2010). Infant Colitis-Iit's in the Genes. *Lancet* 376, 1272. doi:10.1016/s0140-6736(10)61008-2

Glocker, E.-O., Kotlarz, D., Boztug, K., Gertz, E. M., Schäffer, A. A., Noyan, F., et al. (2009). Inflammatory Bowel Disease and Mutations Affecting the Interleukin-10 Receptor. *N. Engl. J. Med.* 361, 2033–2045. doi:10.1056/nejmoa0907206

Griffiths, A. M. (2004). Specificities of Inflammatory Bowel Disease in Childhood. *Best Pract. Res. Clin. Gastroenterol.* 18, 509–523. doi:10.1016/j.bpg.2004.01.002

Heyman, M. B., Kirschner, B. S., Gold, B. D., Ferry, G., Baldassano, R., Cohen, S. A., et al. (2005). Children with Early-Onset Inflammatory Bowel Disease (IBD): Analysis of a Pediatric IBD Consortium Registry. *J. Pediatr.* 146, 35–40. doi:10.1016/j.jpeds.2004.08.043

Horsthuis, K., Bipat, S., Bennink, R. J., and Stoker, J. (2008). Inflammatory Bowel Disease Diagnosed with US, MR, Scintigraphy, and CT: Meta-Analysis of Prospective Studies. *Radiology* 247, 64–79. doi:10.1148/radiol.2471070611

Jardine, S., Anderson, S., Babcock, S., Leung, G., Pan, J., Dhingani, N., et al. (2019). Drug Screen Identifies Leflunomide for Treatment of Inflammatory Bowel Diseases Caused by TTC7A Deficiency. *Gastroenterology* 158 (4), 1000–1015. doi:10.1053/j.gastro.2019.11.019

Jardine, S., Dhingani, N., and Muise, A. M. (2019). TTC7A: Steward of Intestinal Health. *Cell. Mol. Gastroenterol. Hepatol.* 7, 555–570. doi:10.1016/j.jcmgh.2018.12.001

Kammermeier, J., Lucchini, G., Pai, S.-Y., Worth, A., Rampling, D., Amrolia, P., et al. (2016). Stem Cell Transplantation for Tetratricopeptide Repeat Domain 7A Deficiency: Long-Term Follow-Up. *Blood* 128, 1306–1308. doi:10.1182/blood-2016-01-696385

Kelsen, J. R., Russo, P., and Sullivan, K. E. (2019). Early-Onset Inflammatory Bowel Disease. *Immunol. Allergy Clin. N. Am.* 39, 63–79. doi:10.1016/j.iac.2018.08.008

Kelsen, J. R., and Sullivan, K. E. (2017). Inflammatory Bowel Disease in Primary Immunodeficiencies. *Curr. Allergy Asthma Rep.* 17, 57. doi:10.1007/s11882-017-0724-z

Keyashian, K., Dehghan, M., Sceats, L., Kin, C., Limketkai, B. N., and Park, K. T. (2019). Comparative Incidence of Inflammatory Bowel Disease in Different Age Groups in the United States. *Inflamm. Bowel Dis.* 25, 1983–1989. doi:10.1093/ibd/izz092

Lawless, D., Mistry, A., Wood, P. M., Stahlschmidt, J., Arumugakani, G., Hull, M., et al. (2017). Bialellic Mutations in Tetratricopeptide Repeat Domain 7A (TTC7A) Cause Common Variable Immunodeficiency-Like Phenotype with Enteropathy. *J. Clin. Immunol.* 37, 617–622. doi:10.1007/s10875-017-0427-1

Lemoine, R., Pachlopnik-Schmid, J., Farin, H. F., Bigorgne, A., Debré, M., Sepulveda, F., et al. (2014). Immune Deficiency-Related Enteropathy-Lymphocytopenia-Alopecia Syndrome Results from Tetratricopeptide Repeat Domain 7A Deficiency. *J. Allergy Clin. Immunol.* 134, 1354–1364. e6. doi:10.1016/j.jaci.2014.07.019

Lien, R., Lin, Y.-F., Lai, M.-W., Weng, H.-Y., Wu, R.-C., Jaing, T.-H., et al. (2017). Novel Mutations of the Tetratricopeptide Repeat Domain 7A Gene and Phenotype/Genotype Comparison. *Front. Immunol.* 8, 1066. doi:10.3389/fimmu.2017.01066

Matute, J. D., Arias, A. A., Wright, N. A. M., Wrobel, I., Waterhouse, C. C. M., Li, X. J., et al. (2009). A New Genetic Subgroup of Chronic Granulomatous Disease with Autosomal Recessive Mutations in P40 Phox and Selective Defects in Neutrophil NADPH Oxidase Activity. *Blood* 114, 3309–3315. doi:10.1182/blood-2009-07-231498

Muise, A. M., Snapper, S. B., and Kugathasan, S. (2012). The Age of Gene Discovery in Very Early Onset Inflammatory Bowel Disease. *Gastroenterology* 143, 285–288. doi:10.1053/j.gastro.2012.06.025

Muise, A. M., Xu, W., Guo, C.-H., Walters, T. D., Wolters, V. M., Fattouh, R., et al. (2012). NADPH Oxidase Complex and IBD Candidate Gene Studies: Identification of a Rare Variant in NCF2 that Results in Reduced Binding to RAC2. *Gut* 61, 1028–1035. doi:10.1136/gutjnl-2011-300078

Neves, J. F., Afonso, I., Borrego, L., Martins, C., Cordeiro, A. I., Neves, C., et al. (2018). Missense Mutation of TTC7A Mimicking Tricho-Hepato-Enteric (SD/THE) Syndrome in a Patient with Very-Early Onset Inflammatory Bowel Disease. *Eur. J. Med. Genet.* 61, 185–188. doi:10.1016/j.ejmg.2017.11.014

Notarangelo, L. D. (2014). Multiple Intestinal Atresia with Combined Immune Deficiency. *Curr. Opin. Pediatr.* 26, 690–696. doi:10.1097/mop.0000000000000159

Peloquin, J. M., Goel, G., Villablanca, E. J., and Xavier, R. J. (2016). Mechanisms of Pediatric Inflammatory Bowel Disease. *Annu. Rev. Immunol.* 34, 31–64. doi:10.1146/annurev-immunol-032414-112151

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al.ACMG Laboratory Quality Assurance Committee (2015). Standards and Guidelines for the Interpretation of Sequence Variants: a Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* 17, 405–424. doi:10.1038/gim.2015.30

Samuels, M. E., Majewski, J., Alirezaie, N., Fernandez, I., Casals, F., Patey, N., et al. (2013). Exome Sequencing Identifies Mutations in the Gene TTC7A in French-Canadian Cases with Hereditary Multiple Intestinal Atresia. *J. Med. Genet.* 50, 324–329. doi:10.1136/jmedgenet-2012-101483

Shim, J. O. (2019). Recent Advance in Very Early Onset Inflammatory Bowel Disease. *Pediatr. Gastroenterol. Hepatol. Nutr.* 22, 41–49. doi:10.5223/pghn.2019.22.1.41

White, R. A., McNulty, S. G., Nsumu, N. N., Boydston, L. A., Brewer, B. P., and Shimizu, K. (2005). Positional Cloning of the Ttc7 Gene Required for Normal Iron Homeostasis and Mutated in Hea and Fsn Anemia Mice. *Genomics* 85, 330–337. doi:10.1016/j.ygeno.2004.11.008

Worthey, E. A., Mayer, A. N., Syverson, G. D., Helbling, D., Bonacci, B. B., Decker, B., et al. (2011). Making a Definitive Diagnosis: Successful Clinical Application of Whole Exome Sequencing in a Child with Intractable Inflammatory Bowel Disease. *Genet. Med.* 13, 255–262. doi:10.1097/gim.0b013e3182088158

Woutsas, S., Aytekin, C., Salzer, E., Conde, C. D., Apaydin, S., Pichler, H., et al. (2015). Hypomorphic Mutation in TTC7A Causes Combined Immunodeficiency with Mild Structural Intestinal Defects. *Blood* 125, 1674–1676. doi:10.1182/blood-2014-08-595397

Zeytuni, N., and Zarivach, R. (2012). Structural and Functional Discussion of the Tetra-Trico-Peptide Repeat, a Protein Interaction Module. *Structure* 20, 397–405. doi:10.1016/j.str.2012.01.006

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.

Copyright © 2022 Chen, Chen, Guo, Zhang, Li, Xie, Shen, Ge, Huang, Zheng and Lu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 23 June 2022
doi: 10.3389/fgene.2022.921256

m6A Regulator-Mediated RNA Methylation Modification Patterns Regulate the Immune Microenvironment in Osteoarthritis

Yang Duan[1†], *Cheng Yu*[1†], *Meiping Yan*[2], *Yuzhen Ouyang*[3] *and Songjia Ni*[4*]

[1]Department of Spinal Surgery, Zhujiang Hospital, Southern Medical University, Guangdong, China, [2]Outpatient Department, Zhujiang Hospital, Southern Medical University, Guangdong, China, [3]Air Force Hospital of Southern Theater Command of the People's Liberation Army, Guangdong, China, [4]Department of Orthopedics and Traumatology, Zhujiang Hospital, Southern Medical University, Guangdong, China

Epigenetic regulation, particularly RNA n6 methyl adenosine (m6A) modification, plays an important role in the immune response. However, the regulatory role of m6A in the immune microenvironment in osteoarthritis (OA) remains unclear. Accordingly, we systematically studied RNA modification patterns mediated by 23 m6A regulators in 38 samples and discussed the characteristics of the immune microenvironment modified by m6A. Next, we constructed a novel OA m6A nomogram, an m6A-transcription factor-miRNA network, and a drug network. Healthy and OA samples showed distinct m6A regulatory factor expression patterns. *YTHDF3* expression was upregulated in OA samples and positively correlated with type II helper cells and *TGFb* family member receptors. Furthermore, three different RNA modification patterns were mediated by 23 m6A regulatory factors; in Mode 3, the expression levels of *YTHDF3*, type II T helper cells, and *TGFb* family member receptors were upregulated. Pathways related to endoplasmic reticulum oxidative stress and mitochondrial autophagy showed a strong correlation with the regulatory factors associated with Mode 3 and 23 m6A regulatory factors. Through RT-qPCR we validated that *SREBF2* and *EGR1* as transcription factors of *YTHDF3* and *IGF2BP3* are closely associated with the development of OA, hsa-miR-340 as a miRNA for *YTHDF3* and *IGF2BP3* was involved in the development of OA, we also detected the protein expression levels of *IGF2BP3*, *YTHDF3*, *EGR1* and *SREBF2* by western blotting, and the results were consistent with PCR. Overall, the constructed nomogram can facilitate the prediction of OA risk.

Keywords: epigenetics, immune microenvironment, RNA n6 methyl adenosine, osteoarthritis, RNA modification

OPEN ACCESS

Edited by:
Xianyong Yin,
University of Michigan, United States

Reviewed by:
Julong Wei,
Wayne State College, United States
Huan Zhang,
Dana–Farber Cancer Institute, United States

****Correspondence:***
Songjia Ni
warriorkof@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: *21 April 2022*
Accepted: *07 June 2022*
Published: *23 June 2022*

Citation:
Duan Y, Yu C, Yan M, Ouyang Y and Ni S (2022) m6A Regulator-Mediated RNA Methylation Modification Patterns Regulate the Immune Microenvironment in Osteoarthritis. Front. Genet. 13:921256. doi: 10.3389/fgene.2022.921256

INTRODUCTION

Osteoarthritis (OA) is a degenerative multifactorial disease that is characterized by progressive joint failure, which is often associated with joint pain, stiffness, and decreased range of motion (Lopes et al., 2017; Endisha et al., 2021). The pathological changes associated with OA include cartilage degeneration, synovitis, fibrosis, and subchondral bone sclerosis. The etiology of OA is complex and is currently thought to result from a combination of biomechanical processes, trauma, chronic inflammation, and immune response. Previous studies have shown that a variety of cells, cytokines,

chemokines, complements, and other immune system factors are involved in OA pathogenesis (Kandahari et al., 2015). However, studies on the role of m6A regulatory factors in the immune regulation of OA are scarce. Thus, further studies are needed in this regard.

N6 adenosine methylation (M6A), which is the most frequently observed RNA modification type, extensively occurs in mRNA, lncRNAs, and miRNAs. Specifically, M6A plays a crucial role in various physiological processes as well as in disease progression. Its modification is also a dynamic and reversible process controlled by different types of regulatory proteins, namely: methyltransferase ("writer"), demethylase ("erasers"), and the binding protein ("reader"). Furthermore, m6A modification is greatly affected by the expression and function of these regulatory proteins, and studying these regulatory proteins can enhance understanding of the role of m6A in gene regulation (Chong et al., 2021). It has also been observed that m6A modification under the influence of regulatory factors is associated with inflammation, the tumor microenvironment (TME), and immune response (Huang et al., 2021).

Previous studies have shown that m6A regulatory factors, especially *METTL3* and *FTO*, are involved in OA progression via the regulation of inflammatory response and extracellular matrix degradation. On the one hand, FTO-dependent m6A demethylation mediates the upregulation of *AC008*, which inhibits chondrocyte viability and promotes chondrocyte apoptosis and ECM degradation in OA (Yang et al., 2021). On the other hand, *METTL3* affects the stability of autophagy-related 7 (*ATG7*) mRNA, thereby influencing autophagic activity in an m6A- YTHDF2-dependent manner. This in turn promotes FLS fibroblast-like synovial cell senescence and OA progression (Chen et al., 2022). Further, *METTL3* regulates inflammatory responses in OA. It has also been observed that extracellular matrix degradation in OA is related to the balance between *TIMPs* and *MMPs*, which are regulated by *METTL3* (Sang et al., 2021). Furthermore, *METTL3* regulates cartilage tissue by regulating cartilage fine-cellular NF-kB signal transduction, while ECM synthesis plays a mediator role in OA progression (Liu et al., 2019). Moreover, there is increasing evidence that m6A is involved in the regulation of immune responses (Zhang et al., 2021). Therefore, investigating the role of m6A regulatory factors in the immune response of OA and studying the differences in immune changes between healthy and OA tissues can improve our understanding of OA pathogenesis from a completely different perspective.

In this study, we systematically evaluated the modification patterns of m6A regulatory factors in OA. Thus, we found that m6A regulatory factors were distinguished between healthy and OA samples. The abundance of OA-infiltrating immune cells and the immune response genome were found to be significantly correlated with the m6A regulator, thus suggesting the existence of a strong correlation between m6A regulators and immune regulation. We also found different immune characteristics for different m6A molecular subtypes and analyzed their biological functions. These results indicated that the m6A modification pattern has a significant effect on the immune microenvironment in OA. We also constructed an m6A transcription factor-miRNA network as well as a drug network. Simultaneously, a new m6A OA nomogram that can facilitate the prediction of OA risk was established.

MATERIALS AND METHODS

Data Acquisition and Difference Analysis

The relevant OA dataset, GSE114007 (Fisch et al., 2018), was downloaded from the GEO database. This dataset, with data platforms GPL11154 and GPL18573, consists of sequencing data corresponding to Homo sapiens. Further, this GSE114007 dataset includes 38 samples (including 18 control samples and 20 OA samples), all of which were included in this study. The R package DEseq2 (Love et al., 2014) was used to analyze differences in m6A gene expression values between the control and OA groups. The results of the different analyses were presented as heat maps and volcanic maps using R-package heat and ggplot2, respectively.

Analysis of m6A Regulatory Factors in OA

The expression relationships of 23 m6A regulatory factors in healthy and OA samples were evaluated using Spearman's correlation analysis. The random forest (Ishwaran and Lu, 2019) was first used to identify m6A regulators related to OA. Thereafter, the least absolute shrinkage and selection operator (LASSO) regression (Beinse et al., 2022) was used for feature selection, dimensionality reduction, and m6A regulator classifier developments. The distinguishing performances of the signature were then evaluated via receiver operating characteristic (ROC) curve analysis.

Analysis of the Correlation Between m6A Regulators and Immune Characteristics

Single-sample gene set enrichment analysis (ssGSEA) (Hanzelmann et al., 2013) was employed to estimate the number of specific infiltrating immune cells and the activity of specific immune responses. The absolute enrichment degree of a given gene set in each sample within a given dataset was expressed as an enrichment score. Further, gene listings for the gene set of infiltrating immune cells were obtained from previous studies (Shen et al., 2018; Zhang et al., 2020) and the immunoreaction gene set was obtained from the ImmPort database (http://www.immport.org) (Bhattacharya et al., 2014). The enrichment fraction represented the abundance of immune cells as well as the absolute enrichment of the immune response. The correlation of m6A regulators with immunocyte fractions and immune reaction activity was determined using Spearman's correlation analysis.

Identification of m6A Modification Pattern

Based on the expression of 23 m6A regulatory factors, different m6A modification patterns were identified via an unsupervised clustering analysis. Clustering was performed using R-package ConsonsusclusterPlus (Wilkerson and Hayes, 100). The number of clusters and robustness elation map of the patterns were also evaluated. Principal component analysis (PCA) further verified

the expression patterns of the 23 m6A regulators in the different modification modes. Kruskal–Wallis test was then performed to compare the expression of m6A regulatory factors, abundance scores of infiltrating immune cells, immune response scores, and HLA gene expression levels in the three different modification modes.

Biological Enrichment Analysis for Distinct m6A Modification Patterns

Biological signaling pathways may also reflect biological changes. Thus, we obtained the gene sets corresponding to "h.all.v7.0. symbols," and "c5. go.v7.0. symbols" in the MSigDB database and converted the expression matrix into a pathway activation score matrix using the GSVA package. Thereafter, the activation scores of the two groups were compared using the R-package limma (Ritchie et al., 2015). Simultaneously, we also analyzed the correlation between the activation scores corresponding to the endoplasmic reticulum oxidative stress and mitochondrial autophagy pathways and the expression levels of m6A regulatory factors.

Transcription Factors: miRNA Networks and Drug-Compound Networks

We analyzed the targeted transcription factors, targeted miRNAs, and interacting drugs corresponding to high-expression genes in OA using the NetworkAnalyst database (Zhou et al., 2019), and thereafter constructed a network diagram.

Cartilage Donors

Normal human knee cartilage tissues were procured by tissue banks (approved by Scripps Institutional Review Board) from 5 females and 13 males (age 18–61, mean 38) without history of joint disease or trauma and processed within 24–48 h post mortem. Full thickness cartilage was harvested for RNA isolation from identical locations on the weightbearing regions on medial and lateral femoral condyles, and adjacent tissue sections were harvested for histology to verify the cartilage integrity. OA-affected cartilage was harvested from the tissue removed during knee replacement surgery from 12 female and 8 male donors (age 52–82, mean 66). Body mass indices between the normal (BMI = 32.4 ± 8.0) and OA (BMI = 30.7 ± 8.1) were not significantly (p = 0.506) different.

Tissue Processing, RNA and DNA Isolation

Cartilage was stored at −20°C in Allprotect Tissue Reagent (Qiagen, V alencia, CA) immediately after harvest until RNA extraction. For RNA isolation, a minimum of 150 mg of cartilage (dry weight) was pulverized using a 6770 Freezer/Mill Cryogenic Grinder (SPEX SamplePrep, Metuchen, NJ), and homogenized in Qiazol Lysis Reagent (Qiagen, V alencia, CA) at a concentration of 25 mg tissue sample per 700 µl Qiazol. To remove proteins and cellular debris, a initial phenol-chloroform extraction was performed. Briefly, samples were mixed with 0.2 volumes of chloroform, incubated for 5 min in ice, and centrifuged a t 14,000 rpm for 15 min at 4°C. The aqueous phase was collected, mixed with 1 volume of Qiazol and incubated for 30 min in ice. Then, samples were mixed with 1 volume of 100% ethanol, loaded into a mRNeasy Mini kit column (Qiagen) and digested on-column with DNAse following manufacturer instructions. RNA was eluted in 15 µl of RNase-free water. RNA purity was assessed using NanoDrop (ND-1000, Thermo Scientific, Wilmington, DE) and RNA integrity number (RIN) was calculated using a 2100 Bioanalyzer (Agilent, Santa Clara, CA). Average RIN numbers were 6.08 ± 0.95.

Library Preparation and Sequencing

RNA samples from 18 normal and 20 OA cartilage donors were sequenced using 150 ng of total RNA as input. Sequencing mRNA libraries were prepared using the Encore Complete RNA-Seq DR Multiplex System 1–8 and 9–16 (NuGen, San Carlos, CA) with 16 unique indexed adapters (L2V6DR-BC2-L2V6DR-BC16). Two lanes of an Illumina HiSeq 2000 instrument were used to generate a total of 8–30 million 100 bp reads. The Illumina Genome Analyzer Pipeline Software (Casava v1.8.2) was used to convert the original image data generated by the sequencing machine into sequence data via base calling in order to generate fastq files and to demultiplex the samples. We performed a per base sequence quality check using the software FastQC (v0.10.1) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) prior to read mapping. Raw RNAseq reads were aligned to the human genome (hg19) using the STAR aligner. The number of reads sequenced per sample ranged from 19 to 24 million reads, which should be sufficient for gene level quantification, but only 2–12 million reads per sample mapped to protein coding genes. To account for this issue, we applied high stringency the filtering of lowly expressed genes (log counts per million >3) so that only the differential expression analysis included only genes that were expressed in high enough abundances to be confident in their relative gene expression values.

Cell Culture

DMEM, 0.25% trypsin, and phosphate buffered saline (PBS) were equilibrated at room temperature and used for chondrocytes (Procell, Cat NO.: CP-H107, Wuhan, China) culture. Thereafter, the old medium was discarded and washed twice with PBS, which was followed by the addition of an appropriate amount of trypsin and digestion at 37°C for 1–2 min. The digestion was stopped by adding the culture medium, and the chondrocytes were then collected via blowing. The chondrocytes density was adjusted and the cells were inoculated into a 6-well plate, which was placed in an incubator at 37°C with 5% CO_2; the culture was incubated overnight. The next day, the cell culture plate was removed and washed with PBS three times; 1 ml of PBS was added in each wash cycle, and the resulting solution was shaken gently to avoid washing out the cells. Thereafter, the cells in the degeneration group were added to the medium containing IL-1β(PEPROTECH) concentration of 40 ng/ml, while those in the control group were added to the normal medium. Culturing was then continued for 48 h.

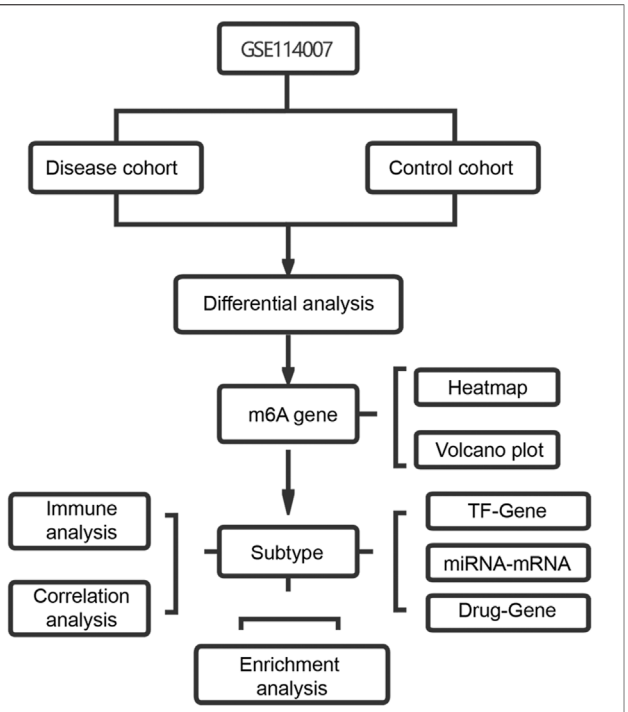


FIGURE 1 | Study flow chart.

Reverse Transcription-Quantitative Polymerase Chain Reaction (RT-qPCR)

Total RNA of 1*106 cells was extracted using TRIzol reagent (15,596–026; Invitrogen, Carlsbad, CA, United States). A reverse transcription reaction system was constructed using the PrimeScript RT reagent kit with GDNA Eraser (# RR047A, Tokyo, Takara), and reverse transcription was performed onboard. The PCR system was constructed using the SYBR Green qPCR Mix (#D7260; Beyotime, Shanghai, China), and PCR detection was performed on a computer system. The 2-ΔΔCt method was used to calculate the relative expression levels in the samples during RT-qPCR data processing. The primers used are listed in **Supplementary Table S1**.

Western Blotting

The total protein of 1*106 cells was extracted using a whole cell lysis assay (KeyGEN Biotech, Nanjing, Jiangsu Province, China). Eighty micrograms of sample protein were subjected to SDS-PAGE (KeyGEN Biotech) and transferred to PVDF membranes (Millipore). The membranes were blocked and probed with the indicated primary antibodies at 4 °C for 12 h. The membranes were then incubated with the indicated HRP-conjugated secondary antibodies at room temperature for 2 h, and the expression of the target proteins was detected by ECL (KeyGEN Biotech). The following antibodies were used: IGF2BP3 (MA5-27484, 1:1000; Thermo Fisher), YTHDF3 (PA5-107309, 1:1000, Thermo Fisher), EGR1 (MA5-15008, 1: 2000; Thermo Fisher), SREBF2 (1:2000; Thermo Fisher) and GAPDH (1:1000; Beyotime Biotechnology, Shanghai, China).

Statistical Analysis

All data calculations and statistical analyses were performed using R software (https://www.r-projec t. org/, version 4.0.2). Comparison between two groups of continuous normally distributed variables were realized by performing the independent Student's t-test, and the difference between non-normally distributed variables was analyzed using the Mann-Whitney U test (Wilcoxon rank-sum test). All statistical p values were bilateral, and statistical significance was set at $p < 0.05$.

RESULTS

Expression of the m6A-Related Genes in OA

Figure 1 shows the flowchart associated with the analysis of the expression of m6A-related genes in OA. To analyze the effect of m6A-related gene expression values on OA tissues relative to normal tissues, differentially expressed m6A-related genes in the dataset were obtained using the DEseq2 package and the results obtained were presented as a volcano plot (**Figure 2A**). From this figure, it is evident that OA showed high *IGF2BP3* and *YTHDF3* expression levels; significantly low *YTHDC1* expression levels; and low *WTAP, IGF2BP2, FMR1, RBM15, ALKBH5, LRPPRC, HNRNPC*, and *METTL14* expression levels. Meanwhile, a heat map (**Figure 2B**), chromosomal circos map (**Figure 2C**), and box map (**Figure 2D**) were generated based on the expression of m6A-related genes. As shown in the abovementioned figures, *YTHDC1, WTAP, IGF2BP2, FMR1, ALKBH5, LRPPRC, HNRNPC*, and *METTL14* showed low expression levels in OA.

Correlation Between the Expression Levels of the m6A-Related Genes in OA

To analyze the correlations between the expression levels of the m6A-related genes in OA, we performed correlation analysis on the expression levels of the m6A-related gene. In this regard, we used the corplot packet to plot the correlation results as a heat map (**Figure 3A**) and network map (**Figure 3B**). **Figures 3C–J** shows the scatter diagrams corresponding to the strongly correlated m6A-related genes in OA. Our results indicated that *METTL3* expression showed a strong correlation with *METTL14* expression, *KIAA1429* expression showed a strong correlation with *YTHDF3* expression, and *RBM15* expression showed a strong correlation with *WTAP* and *LRPPRC* expression. Additionally, *HNRNPA2B1* showed a strong correlation with *RBM15B* and *ALKBH5* expression, and *YTHDC2* showed a strong correlation with *RBM15* and *LRPPRC* expression.

Construction of an m6A Gene Prediction Model in OA

To analyze the resolution of OA by m6A regulatory factors, we first used the random forest method (**Figures 4A,B**). Samples were randomized into a training set (70%) and a verification set (30%), and the boxplots (**Figures 4C,D**) obtained thereafter showed significant differences in model scores between OA and healthy groups in the training and validation sets. Further,

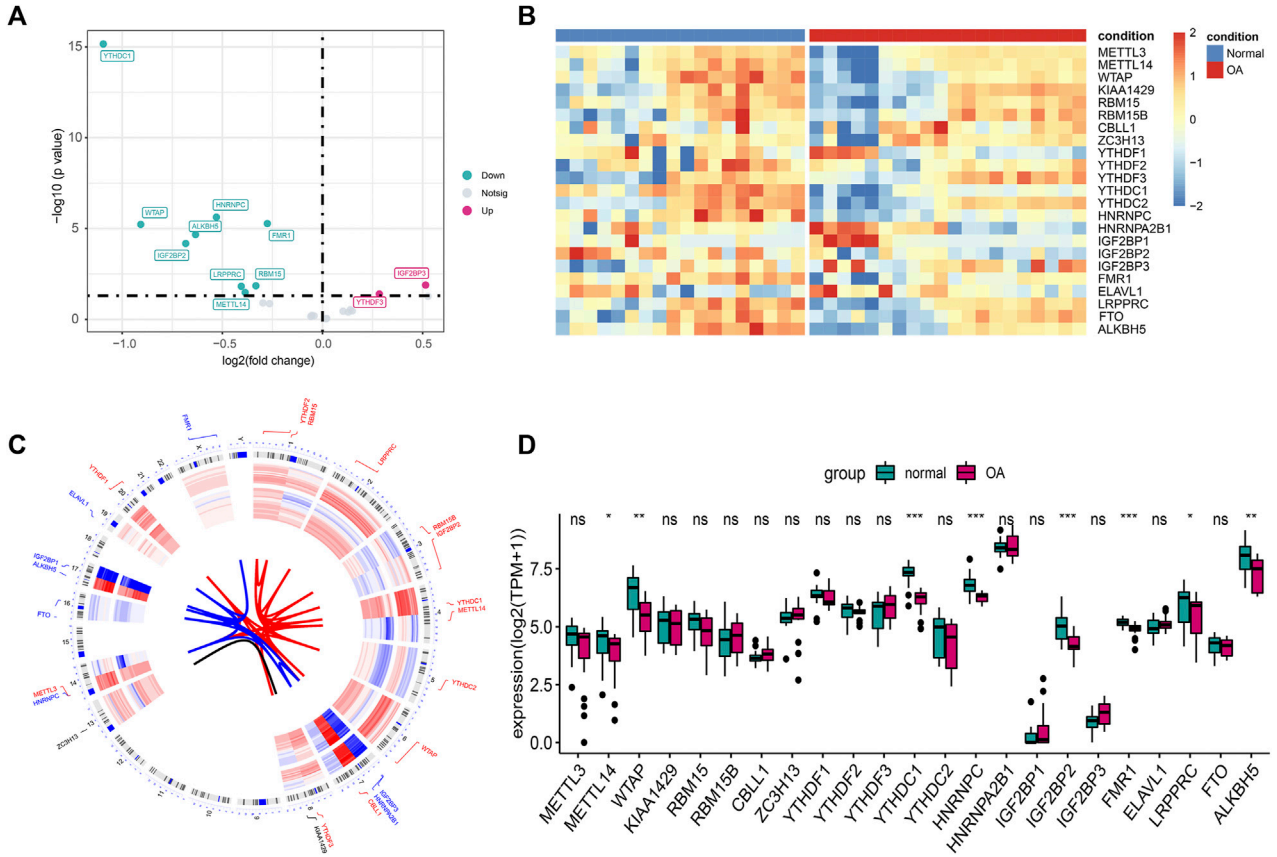


FIGURE 2 | Expression differences of m6A genes in OA. **(A)** The volcano plot of m6A gene difference: the abscissa is log2FC, and the ordinate is −log10(*p*-value). **(B)** Heat map of m6A gene differences. **(C)** Chromosome circus map of differential expression of m6A gene, the inner circle is the expression heat map of OA samples, and the outer circle is the expression heat map of normal samples. **(D)** Boxplot of m6A gene differences. (ns: $p > 0.05$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$).

the ROC curve (**Figure 4E**) showed that the model constructed using the random forest method exhibited a good diagnostic capability for OA; given that the results of the random forest analysis showed that all the m6A regulatory factors had good predictive power, we used Lasso regression analysis to characterize and reduce the m6A regulatory factors while excluding insignificant regulatory factors (**Figures 5A,B**). The obtained prediction model was: Risk Score = *CBLL1* * 2.731 + *ZC3H13* * 7.407 + *YTHDF2* * −0.732 + *YTHDF3* * 2.384 + *YTHDC1* *−12.361 + *IGF2BP2* * −5.030 + *IGF2BP3* * 2.051 + *FMR1* *−5.042. Furthermore, the boxplot in **Figure 5C** showed a significant difference in risk scores between the OA and healthy groups. Subsequently, we constructed an OA risk-related nomogram (**Figure 5D**), which could distinguish between healthy and OA samples as a function of risk scores.

Correlation Between m6A Regulatory Factors in OA and Immune Cells and Immune Process

To investigate the correlation between m6A regulatory factors and the immune microenvironment, we performed a correlation analysis involving m6A regulators, infiltrating immunocytes, and immune reaction gene sets (**Figures 6A,B**). The correlation analysis showed that m6A regulatory factors are strongly correlated with several immune cells. We also found that NK cells were positively correlated with most m6A regulatory factors, while NKT cells and CD8+T cells were negatively correlated with most m6A regulatory factors. Additionally, during the immune process, most of the m6A regulatory factors showed a negative correlation with cytokines and a positive correlation with receptors belonging to the TGFb family.

Modification Patterns of m6A RNA Methylation Mediated by 23 Regulatory Factors in OA

To investigate the m6A modification pattern of bones and joints, we performed unsupervised consensus clustering on OA samples based on the expression of 23 m6A regulatory factors (**Figures 7A–E**). The results thus obtained showed that clustering exhibited good stability at K = 3, and PCA revealed favorable differences among the three m6A molecular subtypes (**Figure 7F**). Further, the thermogram (**Figure 7G**) and boxplot (**Figure 7H**) show that the expression levels of the 23 m6A regulatory factors in the three m6A modification patterns were significantly different.

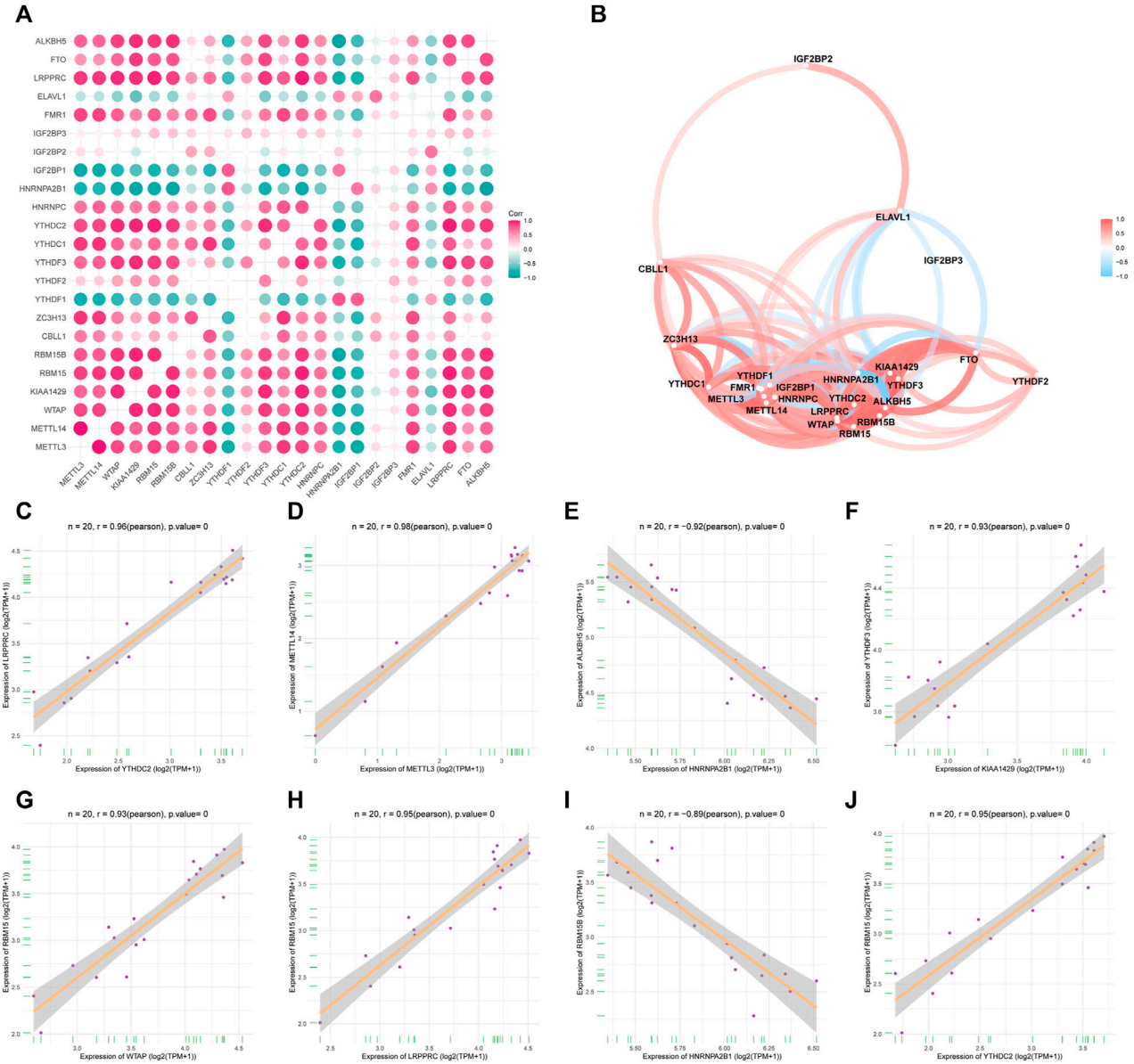


FIGURE 3 | Expression correlation of m6A genes in OA. **(A)** Heatmap of 23 m6A gene expression correlations. **(B)** m6A gene expression correlation network diagram. **(C–J)** Scatter plot of some highly correlated m6A genes.

Immune Microenvironment Characteristics of Different m6A Modification Modes

To understand the differences in the immune microenvironment characteristics of different m6A modification modes, the infiltrating immune cells, immune response genome, and *HLA* gene expression were evaluated. Our results in this regard showed different immune cells for the three patterns (**Supplementary Figure S1A**). Compared to Modes 2 and 3, Mode 1 showed increased immunocytes infiltration; the number of CD4+T cells, immature dendritic cells, and natural killer cells corresponding to Mode 3 were higher than those corresponding to Mode 2. Additionally, members of the TGFB receptor family showed high activity in Mode 3 (**Supplementary Figure S1B**), whereas the cytokine process in Mode 1 was relatively active. This was consistent with the previous analysis of the immune process. Similar trends were observed for *HLA* gene expression (**Supplementary Figure S1C**). In Mode 1, the overall expression of genes belonging to the *HLA* family was high. These results suggested that m6A modifications in Modes 1 and 3 mediated the master immune response, while the modification in Mode 2 mediated the mild immune response to OA. Further, the immune responses mediated by Modes 1 and 3 were different.

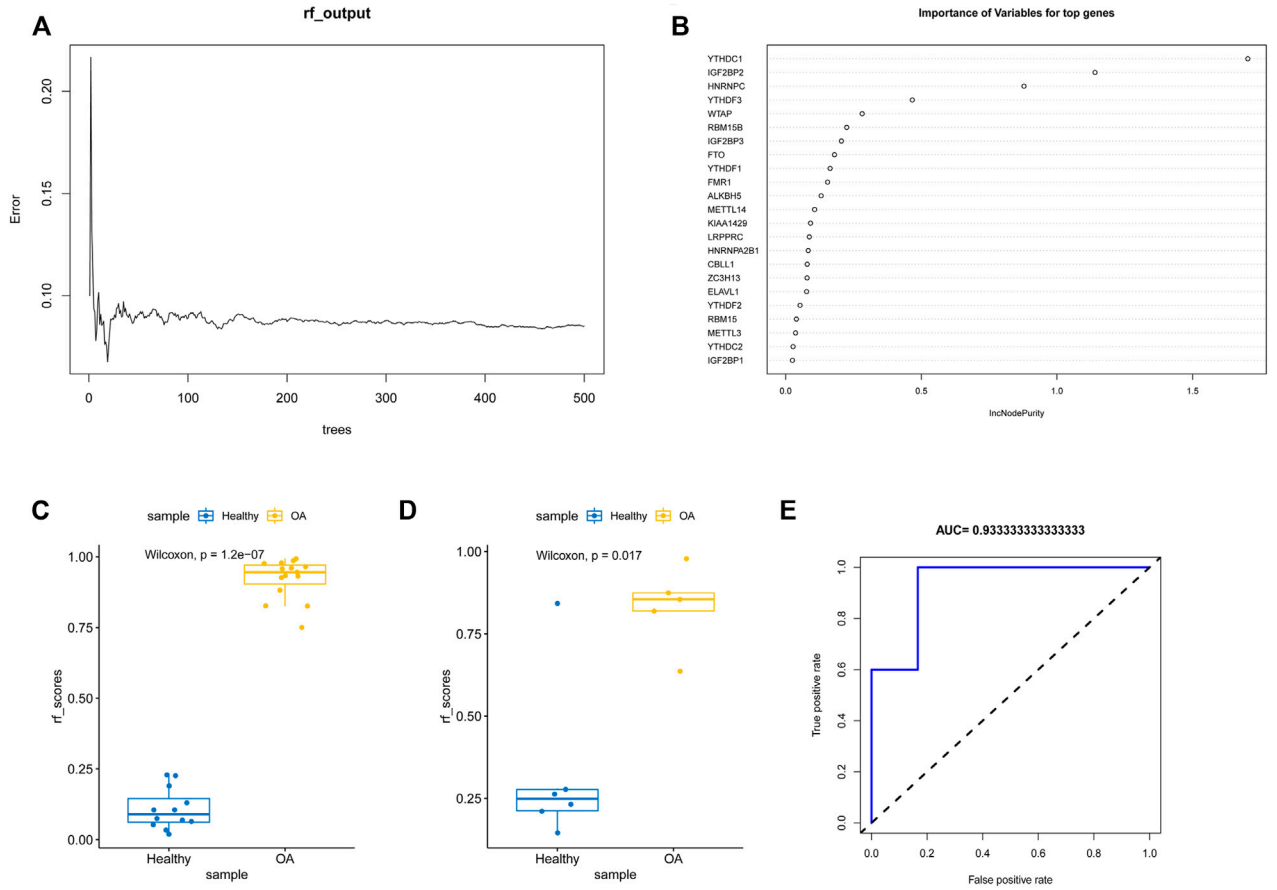


FIGURE 4 | Random forest analysis. **(A,B)** The m6A gene was modeled by random forest. **(C,D)** Boxplots of ratings for training and validation sets. **(E)** ROC plot of random forest diagnosis.

Biological Characteristics of Three m6A Modification Modes

To investigate the biological responses associated with the three m6A modification modes, we compared the HALLMARKS pathway and the GO pathway and evaluated the activation state of the biological pathway using GSVA enrichment analysis (**Supplementary Figures 2A–F**). Compared with Modes 1 and 3, relatively fewer pathways were enriched in Mode 2, thereby revealing the enrichment of pathways such as the ROS pathway. Meanwhile, Modes 1 and 3 were enriched in almost the same pathway. Specifically, Model 1 showed the enrichment of well-known pathways, such as the PI3K, AKT, and MTORC1 pathways, whereas Mode 3 showed the enrichment of pathways related to endoplasmic reticulum oxidative stress and mitochondrial autophagy. Previous studies have demonstrated that endoplasmic reticulum oxidative stress and mitochondrial autophagy play significant roles in OA; therefore, we investigated the correlation between m6A regulatory factors, endoplasmic reticulum oxidative stress, and mitochondrial autophagy pathway scores (**Figure 8A**). These results showed that the m6A regulatory factors were strongly correlated with endoplasmic reticulum oxidative stress and mitochondrial autophagy pathways (**Figures 8B–E**).

m6A-Related TF-MIRNA Network Construction and Drug Development for OA Treatment

In this study, we observed high *IGF2BP3* and *YTHDF3* expression levels in OA. Based on these two genes, we constructed a transcription factor-miRNA network and a drug-compound network diagram using the Network Analyst database. As shown in **Supplementary Figure S3A**, formaldehyde, C646, and other compounds can simultaneously act on these two regulatory factors. This observation suggests that formaldehyde and C646 compounds have the potential to be used as therapeutic drugs. As shown in **Supplementary Figure S3B**, *SREBF1*, *SREBF2*, and *EGR1* were identified as the common targeted transcription factors of the two genes, while hsa-miR-590-3p and hsa-miR-340 were identified as the commonly targeted miRNAs of the two genes.

RT-qPCR and Western Blotting

The expression levels of IGF2BP3 and YTHDF3 were upregulated in the IL-1β-induced degeneration group (**Figures 9A,B**). This is consistent with the results of the

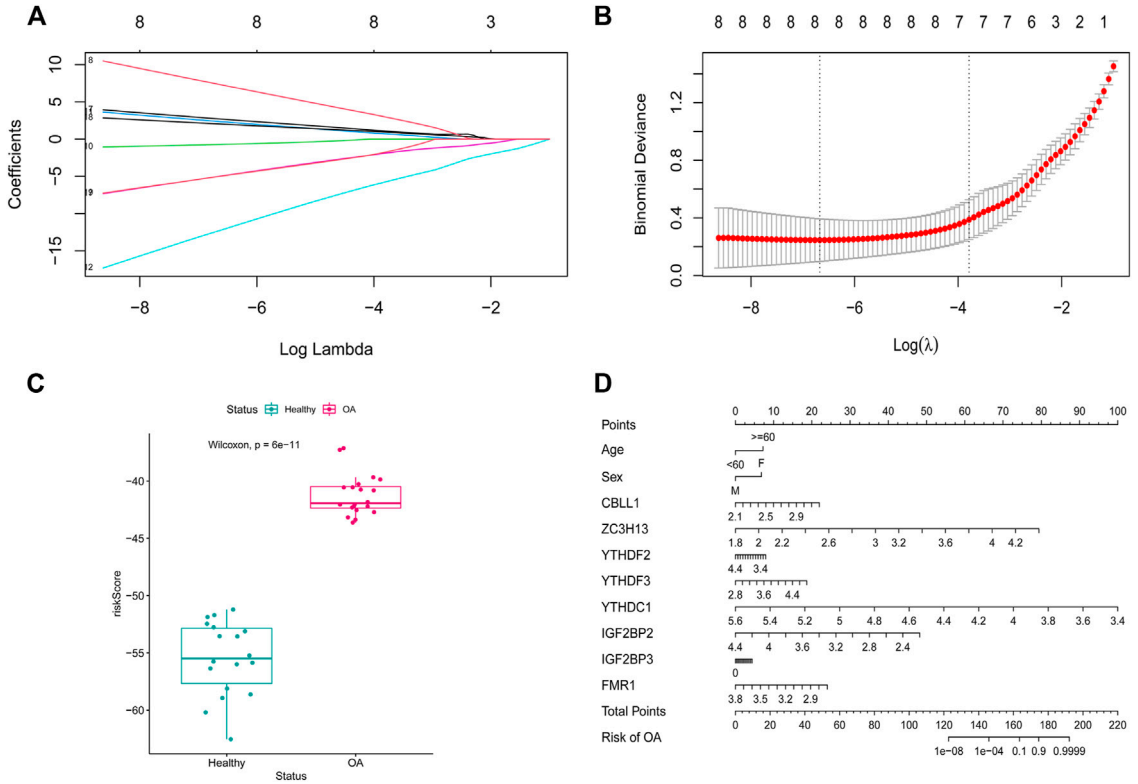


FIGURE 5 | LASSO regression modeling. **(A,B)** LASSO regression modeling m6A gene. **(C)** scoring boxplot. **(D)** diagnostic Nomo plot.

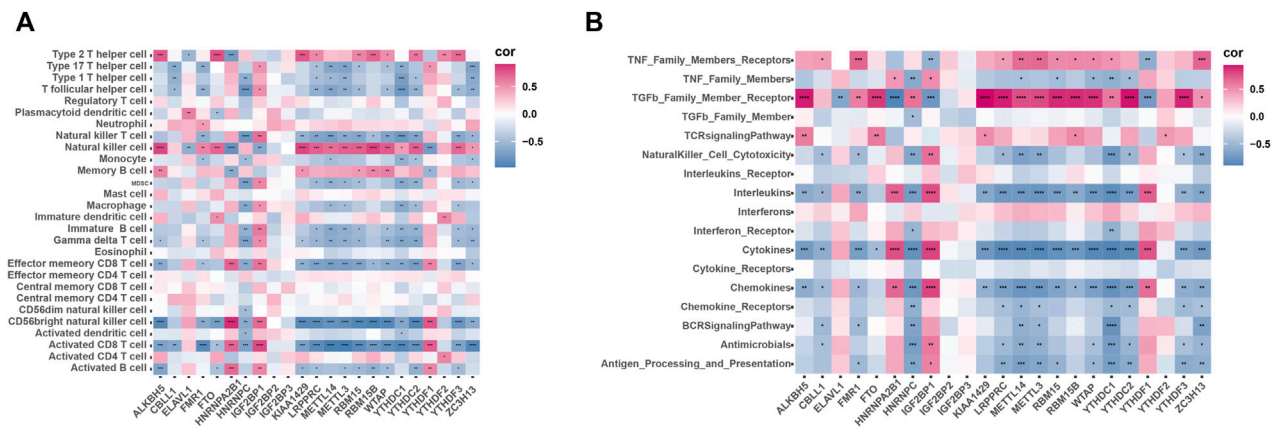


FIGURE 6 | Correlation of infiltrating immune cells, immune response genes and m6A regulators. **(A)** Dot plot shows the correlation between each immune microenvironment aberrant infiltrating cell type and each abnormal m6A regulator. **(B)** Dot plot shows the correlation between each immune response genome and each M6A regulator.

microarray analysis. Further, RT-qPCR results showed the expression of EGR1 and miR-340 of degeneration group was significantly decreased, while the expression of SREBF2 of degeneration group was upregulated (**Figures 9C–E**). These qPCR results for the two groups showed statistically significant differences. On the other hand, we also detected the protein expression levels of IGF2BP3, YTHDF3, EGR1, and SREBF2 by western blotting, and the results were consistent with PCR, IGF2BP3, YTHDF3, and SREBF2 were upregulated in the IL-1β-induced degeneration group, and the expression of EGR1 of degeneration group was significantly decreased. (**Figure 9F**).

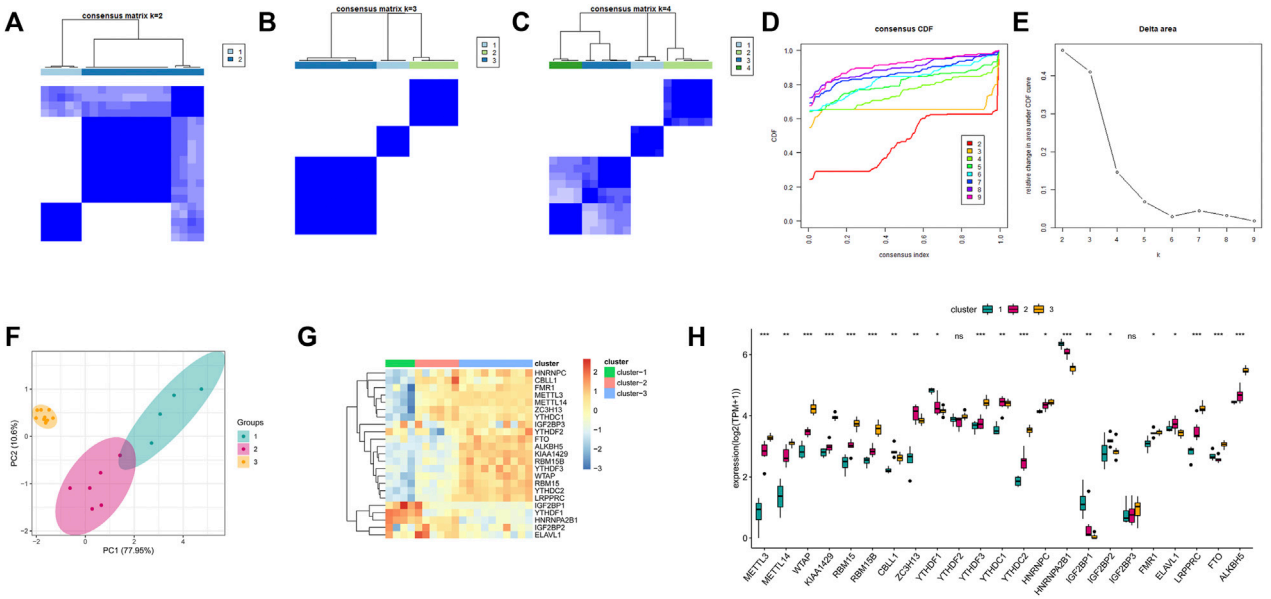


FIGURE 7 | Unsupervised clustering of 23 m6A regulators. Three distinct isoforms of m6A modification patterns were identified in OA. **(A–C)** Heat map of the co-occurrence ratio matrix of OA samples when k = 2–4. **(D)** Consensus clustering cumulative distribution function when k = 2–7. **(E)** Relative area under the CDF curve when k = 2-7 Changes. **(F)** Principal component analysis of the transcriptome profiles of the 3 m6A isoforms showed that the transcriptomes of different modification patterns were significantly different. **(G)** Unsupervised aggregation of 23 m6A regulators in the 3 modification patterns Class. **(H)** Expression status of 23 m6A regulons in 3 m6A isoforms.

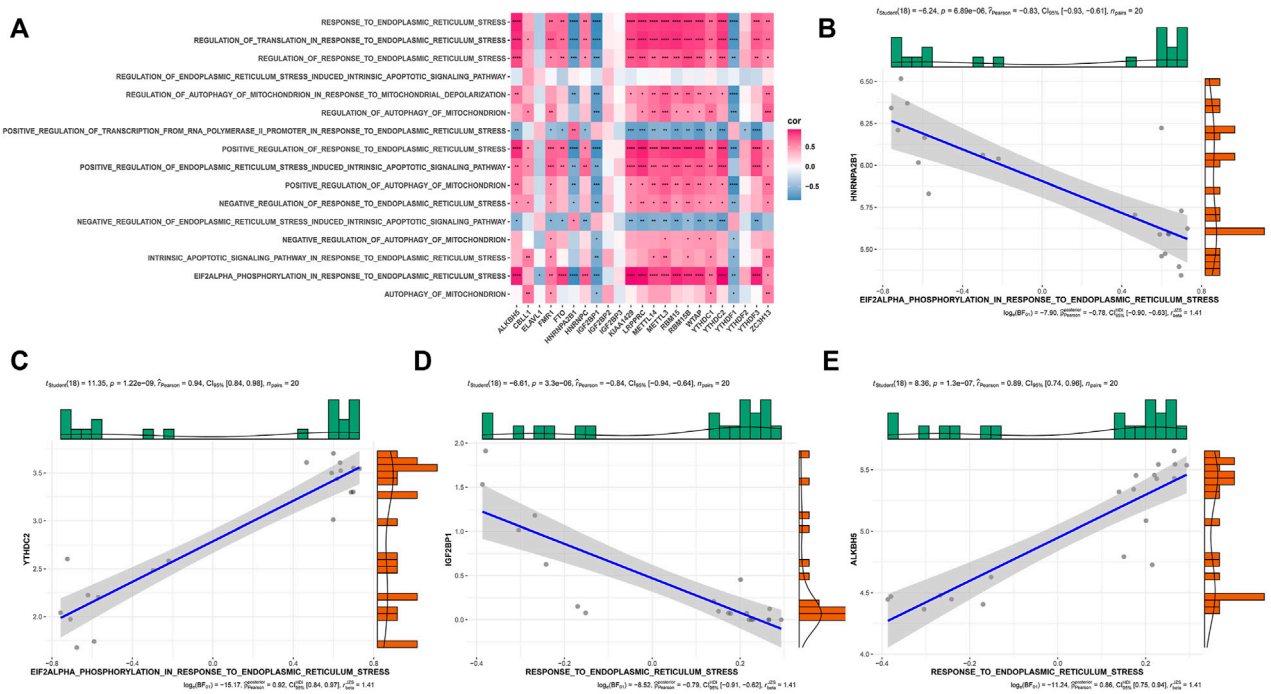


FIGURE 8 | Correlation of m6A gene expression and pathways related to endoplasmic reticulum stress and mitophagy in OA. **(A)** The overall heat map of the correlation between m6A gene expression and ER stress and mitophagy-related pathways. **(B–E)** The correlation between *HNRNPA2B1, YTHDC2, IGF2BP1, and ALKBH5* gene expression and ER stress pathway.

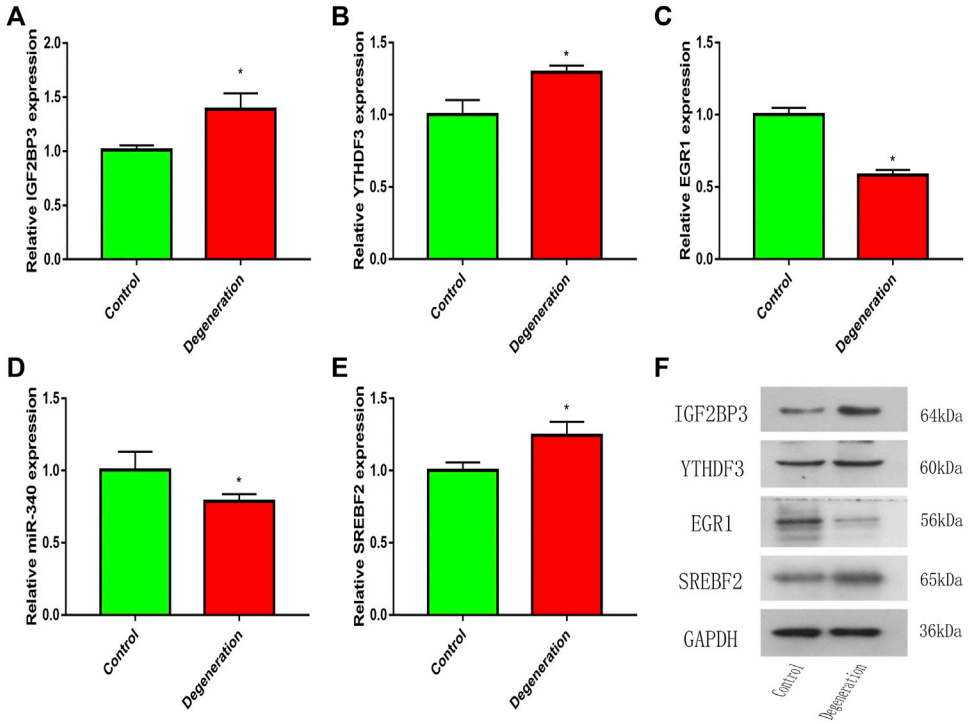


FIGURE 9 | The mRNA or miRNA expression in the control and degeneration group. **(A,B)** Expression level of m6A genes in the control and degeneration group. **(C–E)** Expression level of m6A-related transcription factor and miRNA in the control and degeneration group. **(F)** Expression level of m6A protein and m6A-related transcription factor in the control and degeneration group. *$p < 0.05$ vs. Control.

DISCUSSION

OA is a disease that involves biomechanics, inflammation, and complex biological responses of the immune system (Woodell May and Sommerfeld, 2019). Presently, there are still many research gaps concerning the immune regulatory mechanism of OA, especially m6A regulatory factor-mediated immune regulation. In this study, we systematically investigated the modification pattern of m6A in the OA immune microenvironment; based on the results obtained, we arrived at the following conclusions:

First, we observed a difference in the expression of some m6A regulatory factors between healthy and OA samples, specifically the upregulation of *IGF2BP3* and *YTHDF3* in OA samples. Consistent results were obtained via RT-qPCR and WB, which validated our findings. These results indicated that m6A regulatory factors, especially *IGF2BP3* and *YTHDF3*, may be involved in OA development. Additionally, we observed the existence of correlations between the expression levels of m6A regulatory factors. Thus, we speculated that in OA, m6A regulatory factors jointly regulate OA progression through a regulatory network.

Next, we established an m6A classifier using m6A regulators that offer the possibility to distinguish between healthy and OA samples. This reaffirmed the important role of m6A regulatory factors in OA. Further, to improve and more conveniently predict OA, we established a new m6A nomogram. Based on the review of relevant literature, we found that nomograms are rarely used to predict OA occurrence and progression. Daniel et al. developed a nomogram for diagnosing the rapid progression of knee osteoarthritis. Similarly, we established an m6A nomogram for predicting the risk of OA from the perspective of m6A modifications (Riddle et al., 2016). Different scores were assigned to factors such as age, sex, *CBLL1, ZC3H13, YTHDF2, YTHDF3, YTHDC1, IGF1BP2, IGF2BP3*, and *FMR1*. The total score was obtained by adding the scores of each factor. The total score was less than 160, the probability of OA was less than 0.1, and probability of OA was greater than 0.9 if the total score was greater than 180.

Moreover, we investigated the association between m6A regulatory factors and the immune properties of OA, including the gene set for immune cell infiltration and immune response. We found that many m6A regulatory factors are closely related to these immune characteristics. Unsupervised clustering of periodontitis samples using m6A regulator expression profiles led to three subtypes with distinctive m6A modification patterns, and each subtype exhibited unique immune characteristics. Considering that each subtype has its immune characteristics, we believe that classification based on immunophenotypes of different m6A modulators is feasible. We believe that this classification strategy for immune subtypes will help in comprehensively understanding the mechanisms of immune regulation.

Notably, *YTHDF3* was upregulated in OA samples. In an immune correlation study, *YTHDF3* was positively correlated with type II T helper cells and TGFb family member receptors. In Mode 3, the expression of *YTHDF3*, type II T helper cells, and TGFb family member receptors were upregulated. Mode 3 could be enriched in pathways related to endoplasmic reticulum oxidative stress and mitochondrial autophagy. Meanwhile, we found that the m6A regulatory factor also had a strong correlation with endoplasmic reticulum oxidative stress and mitochondrial autophagy pathway, which is consistent with our previous study. Small extracellular vesicles-miR-151a-3p targeted *YTHDF3* to reduce the transcriptional inhibitory activity of *SP3*, promote the transactivation of TGF-β1 in KCs, and then activate the *SMAD2/3* pathway to enhance the stem cell-like characteristics of the incoming GC cells (Li et al., 2021). TGF-β 1 plays a vital role in maintaining the homeostasis of articular cartilage and subchondral bone (He et al., 2020). Apoptosis of articular chondrocytes is related to ROS-induced oxidative stress, which leads to mitochondrial damage and activates endoplasmic reticulum stress (Feng et al., 2019). Mitochondrial autophagy disorders in chondrocytes accelerate the development of OA (Wang et al., 2020). This suggests that *YTHDF3*, type II helper cells, TGFb family member receptors, endoplasmic reticulum oxidative stress, and mitochondrial autophagy pathways are closely related to OA.

Furthermore, we established a transcription factor-miRNA-m6A regulatory factor network and a drug-compound-m6A regulatory factor network. *SREBF1*, *SREBF2*, and *EGR1* are common transcription factors of *IGF2BP3* and *YTHDF3*, while hsa-miR-590-3p and hsa-miR-340 are two miRNAs that can be combined by both. Compounds such as formaldehyde and *C646* can be combined with these two regulatory factors simultaneously.

SREBP-1 is a transcription factor responsible for the expression of enzymes involved in lipid and cholesterol homeostasis under sterol stimulation. The adsorption of glucose on chitosan membranes (CTS-Glc) stimulated the proliferation of human chondrocytes by providing energy through the regulation of lipogenesis via SREBP-1/Fans and promoting the cell cycle process through the expression of cell cycle regulators induced by *SREBP-1* (Chang et al., 2017).

The cholesterol regulatory element-binding factor-2 (*SREBF2*) gene is a well-known transcriptional regulator of the cholesterol biosynthesis genes. Stigmasterol (*STM*) reduces IL-1β-induced *ATDC5* cell injury in mouse chondrocytes via *SREBF2*, and *STM* reduces il -1β-induced *ATDC5* cell iron ptosis via *SREBF2* (Mo et al., 2021). *SIRT1* may aggravate osteoarthritis cartilage degeneration by activating the *SREBP2* protein-mediated PI3K/AKT signaling pathway (Yu et al., 2016).

EGR1 (early growth response 1) is a transcription factor of the c2h2 type zinc finger protein EGR family that regulates chondrocyte hypertrophy by activating the β-catenin signaling pathway (Sun et al., 2019). The *EGR1* gene has been identified as the central gene of OA development (Chen et al., 2021). *EGR1*, cartilage degeneration, and the expression of *EGR1* in the articular cartilage of OA patients increased (Huan et al., 2021). Mir-340-5p may inhibit the ERK signaling pathway through the *FMOD* gene, promote the proliferation of OA mouse chondrocytes, and inhibit apoptosis (Zhang et al., 2019).

To examine the expression of these genes in healthy and IL-1b-induced osteoarthritis samples, we performed RT-qPCR and confirmed that *IGF2BP3* and *YTHDF3* were closely associated with the development of OA and *SREBF2, EGR1*, and miR-340 could be involved in OA progression by regulating the expression of I*GF2BP3* and *YTHDF3*. However, this study had some limitations. First, the data were downloaded from a public database and could not be evaluated for input errors. Second, based on bioinformatics analysis, RT-qPCR was used to detect the difference in molecular expression between OA and healthy samples; However, flow cytometry still need to be supplemented to verify the role of the molecules and the potential mechanism of OA. Single cells can also be sequenced to obtain the most accurate number of immune cells.

In conclusion, our study revealed a potential regulatory mechanism of m6A modification in the immune microenvironment of OA. Different modification modes of m6A cannot be ignored as they affect the immune microenvironment of OA, thus influencing the occurrence and development of OA. A comprehensive analysis of the modification mode of OA m6A in our study will help to understand the immune regulatory mechanism of OA, provide a reference for the treatment of OA, and supplement the research blank in this field. Meanwhile, the developed m6A OA nomogram can help assess the risk of OA, thus providing a reference for the clinical diagnosis of OA.

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

YD and SN contributed to the study conception and design. Methodology and experiments were performed by YD and CY. Data collection and analysis were performed by MY and YO. The first draft of the manuscript was written by CY. YD and SN revised the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors acknowledge gratitude to all the staff who participated in the Study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.921256/full#supplementary-material

Supplementary Figure S1 | Differences in immune microenvironment characteristics between different m6A modification patterns. **(A)** Differences in the abundance of infiltrating immune cells in each immune microenvironment under the three M6A modification patterns. **(B)** differences in the activity of each immune response genome under the three M6A modification patterns. **(C)** HLA genes in the three m6A modifications Expression differences in patterns.

Supplementary Figure S2 | There are differences in the underlying biological functional characteristics of the three m6A modification patterns. **(A,B)** The difference between the enrichment integrals of M6A modification mode 1 and mode 2 signaling pathways. **(B)** the difference between the enrichment integrals of M6A modification mode 1 and mode 3 signaling pathways. **(C)** M6A modification mode 2 and mode 3 signaling pathways Differences in enrichment scores.

Supplementary Figure S3 | Network graph analysis of highly expressed m6A genes in OA. **(A)** The interaction network of *YTHDF3* and *IGF2BP3* with compounds. **(B)** The transcription factor and miRNA binding network of YTHDF3 and IGF2BP3.

REFERENCES

Beinse, G., Le Frere Belda, M. A., Just, P. A., Bekmezian, N., Koual, M., Garinet, S., et al. (2022). Development and Validation of a RNAseq Signature for Prognostic Stratification in Endometrial Cancer. *Gynecol. Oncol.* 164 (3), 596–606. doi:10.1016/j.ygyno.2022.01.005

Bhattacharya, S., Andorf, S., Gomes, L., Dunn, P., Schaefer, H., Pontius, J., et al. (2014). ImmPort: Disseminating Data to the Public for the Future of Immunology. *Immunol. Res.* 58 (2-3), 234–239. doi:10.1007/s12026-014-8516-1

Chang, S.-F., Huang, K.-C., Cheng, C.-C., Su, Y.-P., Lee, K.-C., Chen, C.-N., et al. (2017). Glucose Adsorption to Chitosan Membranes Increases Proliferation of Human Chondrocyte via Mammalian Target of Rapamycin Complex 1 and Sterol Regulatory Element-Binding Protein-1 Signaling. *J. Cell. Physiol.* 232 (10), 2741–2749. doi:10.1002/jcp.25869

Chen, H., Ni, Q., Li, B., and Chen, L. (2021). Identification of Differentially Expressed Genes in Synovial Tissue of Osteoarthritis Based on a More Robust Integrative Analysis Method. *Clin. Rheumatol.* 40 (9), 3745–3754. doi:10.1007/s10067-021-05649-z

Chen, X., Gong, W., Shao, X., Shi, T., Zhang, L., Dong, J., et al. (2022). METTL3-mediated m6A Modification of ATG7 Regulates Autophagy-GATA4 axis to Promote Cellular Senescence and Osteoarthritis Progression. *Ann. Rheum. Dis.* 81 (1), 87–99. doi:10.1136/annrheumdis-2021-221091

Chong, W., Shang, L., Liu, J., Fang, Z., Du, F., Wu, H., et al. (2021). m6A Regulator-Based Methylation Modification Patterns Characterized by Distinct Tumor Microenvironment Immune Profiles in Colon Cancer. *Theranostics* 11 (5), 2201–2217. doi:10.7150/thno.52717

Endisha, H., Datta, P., Sharma, A., Nakamura, S., Rossomacha, E., Younan, C., et al. (2021). MicroRNA-34a-5p Promotes Joint Destruction during Osteoarthritis. *Arthritis Rheumatol.* 73 (3), 426–439. doi:10.1002/art.41552

Feng, K., Chen, Z., Pengcheng, L., Zhang, S., and Wang, X. (2019). Quercetin Attenuates Oxidative Stress-induced Apoptosis via SIRT1/AMPK-mediated Inhibition of ER Stress in Rat Chondrocytes and Prevents the Progression of Osteoarthritis in a Rat Model. *J. Cell. Physiology* 234 (10), 18192–18205. doi:10.1002/jcp.28452

Fisch, K. M., Gamini, R., Alvarez-Garcia, O., Akagi, R., Saito, M., Muramatsu, Y., et al. (2018). Identification of Transcription Factors Responsible for Dysregulated Networks in Human Osteoarthritis Cartilage by Global Gene Expression Analysis. *Osteoarthr. Cartil.* 26 (11), 1531–1538. doi:10.1016/j.joca.2018.07.012

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* 33 (1), 1–22.

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC Bioinforma.* 14 (1), 7. doi:10.1186/1471-2105-14-7

He, J., Cao, W., Azeem, I., and Shao, Z. (2020). Epigenetics of Osteoarthritis: Histones and TGF-B1. *Clin. Chim. Acta* 510, 593–598. doi:10.1016/j.cca.2020.08.011

Huan, Z., Wang, Y., Zhang, M., Zhang, X., Liu, Y., Kong, L., et al. (2021). Follicle-stimulating Hormone Worsens Osteoarthritis by Causing Inflammation and Chondrocyte Dedifferentiation. *FEBS Open Bio* 11 (8), 2292–2303. doi:10.1002/2211-5463.13238

Huang, X., Qiu, Z., Li, L., Chen, B., and Huang, P. (2021). m6A Regulator-Mediated Methylation Modification Patterns and Tumor Microenvironment Infiltration Characterization in Hepatocellular Carcinoma. *Aging* 13 (16), 20698–20715. doi:10.18632/aging.203456

Ishwaran, H., and Lu, M. (2019). Standard Errors and Confidence Intervals for Variable Importance in Random Forest Regression, Classification, and Survival. *Statistics Med.* 38 (4), 558–582. doi:10.1002/sim.7803

Kandahari, A. M., Yang, X., Dighe, A. S., Pan, D., and Cui, Q. (2015). Recognition of Immune Response for the Early Diagnosis and Treatment of Osteoarthritis. *J. Immunol. Res.* 2015, 192415. doi:10.1155/2015/192415

Li, B., Xia, Y., Lv, J., Wang, W., Xuan, Z., Chen, C., et al. (2021). miR-151a-3p-rich Small Extracellular Vesicles Derived from Gastric Cancer Accelerate Liver Metastasis via Initiating a Hepatic Stemness-Enhancing Niche. *Oncogene* 40 (43), 6180–6194. doi:10.1038/s41388-021-02011-0

Liu, Q., Li, M., Jiang, L., Jiang, R., and Fu, B. (2019). METTL3 Promotes Experimental Osteoarthritis Development by Regulating Inflammatory Response and Apoptosis in Chondrocyte. *Biochem. Biophys. Res. Commun.* 516 (1), 22–27. doi:10.1016/j.bbrc.2019.05.168

Lopes, E. B. P., Filiberti, A., Husain, S. A., and Humphrey, M. B. (2017). Immune Contributions to Osteoarthritis. *Curr. Osteoporos. Rep.* 15 (6), 593–600. doi:10.1007/s11914-017-0411-y

Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8

Mo, Z., Xu, P., and Li, H. (2021). Stigmasterol Alleviates Interleukin-1beta-Induced Chondrocyte Injury by Down-Regulatingsterol Regulatory Element Binding Transcription Factor 2 to Regulateferroptosis. *Bioengineered* 12 (2), 9332–9340. doi:10.1080/21655979.2021.2000742

Riddle, D. L., Stratford, P. W., and Perera, R. A. (2016). The Incident Tibiofemoral Osteoarthritis with Rapid Progression Phenotype: Development and Validation of a Prognostic Prediction Rule. *Osteoarthr. Cartil.* 24 (12), 2100–2107. doi:10.1016/j.joca.2016.06.021

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

Sang, W., Xue, S., Jiang, Y., Lu, H., Zhu, L., Wang, C., et al. (2021). METTL3 Involves the Progression of Osteoarthritis Probably by Affecting ECM Degradation and Regulating the Inflammatory Response. *Life Sci.* 278, 119528. doi:10.1016/j.lfs.2021.119528

Shen, S., Wang, G., Zhang, R., Zhao, Y., Yu, H., Wei, Y., et al. (2018). Development and Validation of an Immune Gene-Set Based Prognostic Signature in Ovarian Cancer. *EBioMedicine* 40, 318–326. doi:10.1016/j.ebiom.2018.12.054

Sun, X., Huang, H., Pan, X., Li, S., Xie, Z., Ma, Y., et al. (2019). EGR1 Promotes the Cartilage Degeneration and Hypertrophy by Activating the Krüppel-like Factor 5 and β-catenin Signaling. *Biochim. Biophys. Acta (BBA) - Mol. Basis Dis.* 1865 (9), 2490–2503. doi:10.1016/j.bbadis.2019.06.010

Wang, F.-S., Kuo, C.-W., Ko, J.-Y., Chen, Y.-S., Wang, S.-Y., Ke, H.-J., et al. (2020). Irisin Mitigates Oxidative Stress, Chondrocyte Dysfunction and Osteoarthritis Development through Regulating Mitochondrial Integrity and Autophagy. *Antioxidants* 9 (9), 810. doi:10.3390/antiox9090810

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: a Class Discovery Tool with Confidence Assessments and Item Tracking. *Bioinformatics* 26, 1572–1573. doi:10.1093/bioinformatics/btq170

Woodell may, J. E., and Sommerfeld, S. D. (2019). Role of Inflammation and the Immune System in the Progression of Osteoarthritis. *J. Orthop. Res.* 38 (2), 253–257. doi:10.1002/jor.24457

Yang, J., Zhang, M., Yang, D., Ma, Y., Tang, Y., Xing, M., et al. (2021). m6A-mediated Upregulation of AC008 Promotes Osteoarthritis Progression through the miR-328-3p–AQP1/ANKH axis. *Exp. Mol. Med.* 53 (11), 1723–1734. doi:10.1038/s12276-021-00696-7

Yu, F., Zeng, H., Lei, M., Xiao, D.-m., Li, W., Yuan, H., et al. (2016). Effects of SIRT1 Gene Knock-Out via Activation of SREBP2 Protein-Mediated PI3K/AKT Signaling on Osteoarthritis in Mice. *J. Huazhong Univ. Sci. Technol. Med. Sci.]* 36 (5), 683–690. doi:10.1007/s11596-016-1645-0

Zhang, W., Cheng, P., Hu, W., Yin, W., Guo, F., Chen, A., et al. (2019). Downregulated microRNA-340-5p Promotes Proliferation and Inhibits Apoptosis of Chondrocytes in Osteoarthritis Mice through Inhibiting the Extracellular Signal-regulated Kinase Signaling Pathway by Negatively Targeting the FMOD Gene. *J. Cell. Physiol.* 234 (1), 927–939. doi:10.1002/jcp.26921

Zhang, B., Wu, Q., Li, B., Wang, D., Wang, L., and Zhou, Y. L. (2020). m6A Regulator-Mediated Methylation Modification Patterns and Tumor Microenvironment Infiltration Characterization in Gastric cancerA Regulator-Mediated Methylation Modification Patterns and Tumor Microenvironment Infiltration Characterization in Gastric Cancer. *Mol. Cancer* 19 (1), 53. doi:10.1186/s12943-020-01170-0

Zhang, X., Zhang, S., Yan, X., Shan, Y., Liu, L., Zhou, J., et al. (2021). m6A Regulator-mediated RNA Methylation Modification Patterns Are Involved in Immune Microenvironment Regulation of Periodontitis. *J. Cell. Mol. Med.* 25 (7), 3634–3645. doi:10.1111/jcmm.16469

Zhou, G., Othman, S., Jessica, E., Hancock, R., Niladri, B., and Xia, J. (2019). NetworkAnalyst 3.0: a Visual Analytics Platform for Comprehensive Gene Expression Profiling and Meta-Analysis. *Nucleic Acids Res.* 47, W234–W241. doi:10.1093/nar/gkz240

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.

Copyright © 2022 Duan, Yu, Yan, Ouyang and Ni. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 10 August 2022
DOI 10.3389/fgene.2022.891465

Check for updates

OPEN ACCESS

EDITED BY
Sukhbir Kaur,
National Institutes of Health (NIH),
United States

REVIEWED BY
Jeyakumar Natarajan,
Bharathiar University, India
Sammed Mandape,
University of North Texas Health
Science Center, United States

*CORRESPONDENCE
Qing Sun,
sunqing7226@163.com

SPECIALTY SECTION
This article was submitted to Genetics of
Common and Rare Diseases,
a section of the journal
Frontiers in Genetics

RECEIVED 07 March 2022
ACCEPTED 18 July 2022
PUBLISHED 10 August 2022

CITATION
Yan J, Zhen Y, Wang R, Li X, Huang S,
Zhong H, Wen H and Sun Q (2022),
Downregulation of miRNA miR-1305
and upregulation of miRNA miR-6785-
5p may be associated with psoriasis.
Front. Genet. 13:891465.
doi: 10.3389/fgene.2022.891465

COPYRIGHT
© 2022 Yan, Zhen, Wang, Li, Huang,
Zhong, Wen and Sun. This is an open-
access article distributed under the
terms of the Creative Commons
Attribution License (CC BY). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Downregulation of miRNA miR-1305 and upregulation of miRNA miR-6785-5p may be associated with psoriasis

Jianjun Yan[1,2], Yunyue Zhen[1,2], Ruijie Wang[1,2], Xueqing Li[1,2], Shan Huang[1,2], Hua Zhong[1,2], He Wen[1,2] and Qing Sun[1,2]*

[1]Department of Dermatology, Qilu Hospital of Shandong University, Jinan, China, [2]Laboratory of Basic Medical Science, Qilu Hospital of Shandong University, Jinan, China

Background: The role of serum extracellular vesicles (EVs) is less known in psoriasis.

Objectives: To explore the transcriptomic profile of serum EVs and the potential biomarkers in psoriasis.

Methods: EVs were isolated by differential ultracentrifugation and identified by transmission electron microscope. The diameters of EVs were detected using nanoparticle tracking analysis. Serum EVs-keratinocyte interaction was observed through confocal fluorescence microscopy. miRNA microarray and mRNA microarray were performed in serum EVs (n = 4) and skin lesions (n = 3), respectively. Quantitative reverse-transcriptase polymerase chain reaction (qRT-PCR) and fluorescence *in situ* hybridization were used to detect the expression of miRNAs in serum EVs and skin lesions (n = 15). Bioinformatics analysis was performed to predict the potential target genes and functions of miR-1305 and miR-6785-5p. Western blot, CCK-8 and enzyme-linked immunosorbent assay (ELISA) were used to detect the EVs' biomarkers, keratinocytes proliferation and cytokines secretion.

Results: A total of 16 miRNAs and 1,725 mRNAs were significantly dysregulated in serum EVs and skin lesions, respectively. miR-1305 was down-regulated and miR-6785-5p was upregulated in both serum EVs and skin lesions. Serum EVs could be taken up by keratinocytes. miR-1305 was downregulated and miR-6785-5p were upregulated in keratinocytes after co-cultured with psoriasis serum EVs compared with controls. Psoriasis serum EVs promoted keratinocyte proliferation and the secretion of CCL20 and IL-8. Serum EVs miR-1305 and miR-6785-5p were associated with disease severity.

Conclusion: Serum EVs might be involved in the activation of keratinocytes through loaded miRNAs in psoriasis. Serum EVs miR-1305 and miR-6785-5p may be associated with psoriasis.

KEYWORDS

extracellular vesicles, miR-1305, MiR-6785-5p, psoriasis, inflammation

Introduction

Psoriasis is a chronic inflammatory skin disease (Srivastava et al., 2017; Nussbaum et al., 2021). The worldwide prevalence rate of this disease is about 2%–3% (Boehncke and Schon, 2015). Multiple factors contribute to the skin inflammation of psoriasis, including genetic, environmental, and immunologic triggers (Frischknecht et al., 2019; Griffiths et al., 2021). Psoriasis was proposed to result from a complex interplay among keratinocytes, immune cells and inflammatory mediators (Guinea-Viniegra et al., 2014). Keratinocytes are the main constituents of the epidermis and play an important role in the formation of psoriatic skin inflammation (Ni and Lai, 2020; Chen et al., 2021). Previous studies have suggested that the intercellular communication between keratinocytes and immune cells is mediated by proinflammatory cytokines (Than et al., 2019; Papayannakos et al., 2021). CCL20 and IL-8 secreted by keratinocytes play important roles in psoriasis, including the activation and chemotaxis of Th17 cells and neutrophils (Tsai and Tsai, 2017; Furue et al., 2020).

Recently, extracellular vesicles (EVs) are emerging mediators of intercellular communication between keratinocytes and immune cells, thought to participate in the pathogenesis of psoriasis (Shao et al., 2020). EVs are secreted from numerous cell types and have been isolated from a wide variety of human body fluids such as blood, urine, and saliva (Kim et al., 2017). EVs represent an important mode of intercellular communication by serving as vehicles for transfer between cells of membrane and cytosolic proteins, lipids, and RNA. Its functions depend on the ability of EVs to interact with recipient cells to deliver their contents (Raposo and Stoorvogel, 2013). A previous study has shown that keratinocyte-derived EVs promoted the Th1/Th17 polarization in psoriasis (Jiang et al., 2021). However, the characterization, transcriptomic profile and functional role of serum EVs in mediating keratinocyte activation and psoriatic skin inflammation are yet to be explored.

MicroRNAs (miRNAs) are 18–22-nucleotide, single-stranded, endogenous noncoding small RNA molecules. MiRNAs target mRNAs through complementary base pairing with 3′-UTRs of the target transcripts, in either a complete or incomplete fashion (Lu and Rothenberg, 2018). MiRNAs function by inhibiting mRNA translation and stability. Recently, miRNA dysregulation in the pathogenesis of psoriasis has attracted attention (Wu et al., 2018). However, little is known about the transcriptomic profile of miRNAs in serum EVs and the role of serum EVs miRNAs in the activation of keratinocytes in psoriasis.

Previous studies have confirmed the intrinsic cell targeting properties of EVs (Vader et al., 2016). However, recent studies mainly focused on the communication between EVs and neighbouring cells in psoriasis, such as T cells or neutrophils (Cheung et al., 2016; Shao et al., 2019). The communication between serum EVs and keratinocytes remains less known. As the cell-targeting properties of EVs and the serum EVs-keratinocyte interaction were confirmed in this study, we assume that serum EVs might not only communicate with neighbouring cells but also communicate with keratinocytes in psoriasis.

In this study, we detected the characterization of serum EVs from psoriasis patients and healthy controls, and verified serum EVs could be taken up by keratinocytes through confocal fluorescence microscopy (CFM). To investigate the regulatory role of serum EVs in the activation of keratinocytes, we isolated serum EVs and performed a miRNA microarray. The mRNA microarray was also performed in skin lesions. In addition, we verified the role of serum EVs and the expression of serum EVs miR-1305 and miR-6785-5p. Previous studies suggest that miR-1305 and miR-6785-5p play important roles in diseases by regulating TGF-β2/smad3 pathway, Wnt/beta-catenin pathway, Akt-signaling pathway and so on (Yu et al., 2019; Liu et al., 2020; Su et al., 2020; Kim et al., 2021). Furthermore, we found miR-1305 and miR-6785-5p might be involved in regulating inflammation response and correlated with the disease severity of psoriasis.

Our findings indicated a potential role of serum EVs in the epidermal hyperplasia and skin inflammation in psoriasis, opening the possibility that the serum EVs miRNAs might be as biomarkers and that serum EVs-keratinocytes interaction can be harnessed as a platform for innovative therapeutic strategies in psoriasis.

Materials and methods

Patients and tissue samples

Fifteen specimens were obtained from patients with psoriasis vulgaris (ten males and five females, aged 25–59 years). They had received no systemic treatments, phototherapy or externally used drugs for at least 1 month before skin biopsies and blood specimens collected. Ten control skin biopsies were obtained from healthy volunteers (five males and five females, aged 22–57 years). We obtained biopsies (1 × 0.5 cm) by surgical operations and whole blood (5–10 ml). Psoriasis Area Severity Index (PASI) was determined at the condition when samples were collected by at least two dermatologists. The study was approved by the ethics committee of Shandong University, China, and all patients provided written informed consent.

Serum extracellular vesicles extraction

Whole blood (5–10 ml) from patients and healthy controls were centrifuged at 3,000 rpm for 10 min to separate serum. The serum was collected, centrifuged at 300 g for 10 min, then at 16,500 g for 30 min, and later filtered through a 0.22 mm pore size filter to remove cell debris. The final supernatant was

ultracentrifuged at 100,000 g for 70 min and recentrifuged at the same speed. The purified EVs were resuspended in PBS.

Characterization of serum extracellular vesicles

Morphology of serum EVs was observed by transmission electron microscopy (TEM). The numbers and diameters were detected using nanoparticle tracking analysis (NTA, Zeta-View PMX 110; Particle Metrix, Meerbusch, Germany). Phenotypic was assessed by western blot using antibodies (Abcam) specific for CD9 (1:2,000, ab92726), CD63 (1:1,000, ab134045), and CD81 (1:1,000, ab109201).

miRNA and mRNA microarray analysis

Serum EVs miRNAs and mRNAs in skin lesions were detected using Agilent Human miRNA microarray V 21.0 (Sinotech Genomics, Shenzhen, China) and SBC Human (4 × 180K) competing endogenous (ce)RNA microarray (Shanghai Biotechnology Corporation, Shanghai, China), respectively. The fold change >2 in these analyses was an absolute value of fold change.

Fluorescence *in situ* hybridization analysis

A diluted double CY3-labelled miR-1305 probe (8 ng/μl, 5′-CY3-AAAAGTTGAGATTACCCTCTCT-CY3-3’; Servicebio, Wuhan, China) and miR-6785-5p probe (8 ng/μl, 5′-CY3-ACCCTCCCGCACCTACTACCAC-CY3-3’; Servicebio, Wuhan, China) were used for hybridization. 4′,6-Diamidino-2-phenylindole (Servicebio, Wuhan, China) was used for nuclear staining.

Serum extracellular vesicles-keratinocytes interaction experiment

Serum EVs were labelled with PKH26 Red Fluorescent Cell Linker Mini kit (MINI26-1 KT; Sigma-Aldrich, St. Louis, MO) according to the instructions. Then, the suspension was centrifuged at 100,000 g for 70 min, and the supernatant was discarded. The labelled EVs were resuspended in PBS and centrifuged once again at the same speed. Next, the EVs were co-cultured with normal human epidermal keratinocytes (NHEKs) cells for 12 h. Then washed the cells 3 times with PBS and fixed with 4% formaldehyde. Subsequently, the cells were stained using phalloidin-iFluor 488 reagent, and DAPI was added to stain the nuclei. The cells were then observed under a confocal microscope (LSM800; Zeiss).

Quantitative reverse-transcriptase polymerase chain reaction

Serum EVs were digested with RNaseA (100 ng/ml, Qiagen, Australia). Then, total RNA from serum EVs was extracted using TRIzol reagent (Invitrogen). The expression levels of serum EVs miRNAs and miRNAs in skin lesions were detected according to the manual of the All-in-One miRNA qRT-PCR Detection System (GeneCopoeia, Guangzhou, China). U6 was used as the internal control for miRNA. Total RNAs from skin samples were extracted using an RNeasy mini kit (cat. no. 74106; Qiagen GmBH, Hilden, Germany) according to the manufacturer’s protocols. Then the expression levels of mRNAs were detected using microarray analysis.

Western blot analysis

Western blot was performed as described in our previous experiments (Jiang et al., 2017; Zhao et al., 2017). The following primary antibodies were used in this study: rabbit monoclonal antibodies (procured from Abcam) specific for CD9 (1:2,000, ab92726), CD63 (1:1,000, ab134045) and CD81 (1:1,000, ab109201) and GAPDH (1:1,000, ab8245).

Bioinformatics analysis and miRNA-mRNA net-work analysis

Target genes of miRNAs were predicted using TargetScanHuman 7.1 and miRDB-MicroRNA Target Prediction Database. Then, the genes were intersected with mRNAs detected by mRNA microarray. Enrichment calculations were performed using Fisher’s exact test. Further, we conducted GO and pathway enrichment analysis of the target genes. The specific principle was to carry out annotation mapping of differentially expressed genes in GO and KEGG database entries, calculate the number of the target genes in each GO and pathway entry, and then use hypergeometric test for statistics. Select the GO and KEGG entries that were significantly enriched in the differentially expressed genes. After the calculated *p*-value was corrected by multiple hypothesis tests, the *p*-value 0. 05 was taken as the threshold, and the GO and KEGG term meeting this condition was defined as the GO and KEGG term significantly enriched in the target genes. MiRNA-mRNA network analysis was visualized using Cytoscape 3.9.0 software.

CCK-8 proliferation assay

The proliferation rates of NHEK Cells were measured at different time intervals using CCK-8 assays (Beyotime, Shanghai, China) after co-cultured with serum EVs obtained from psoriasis

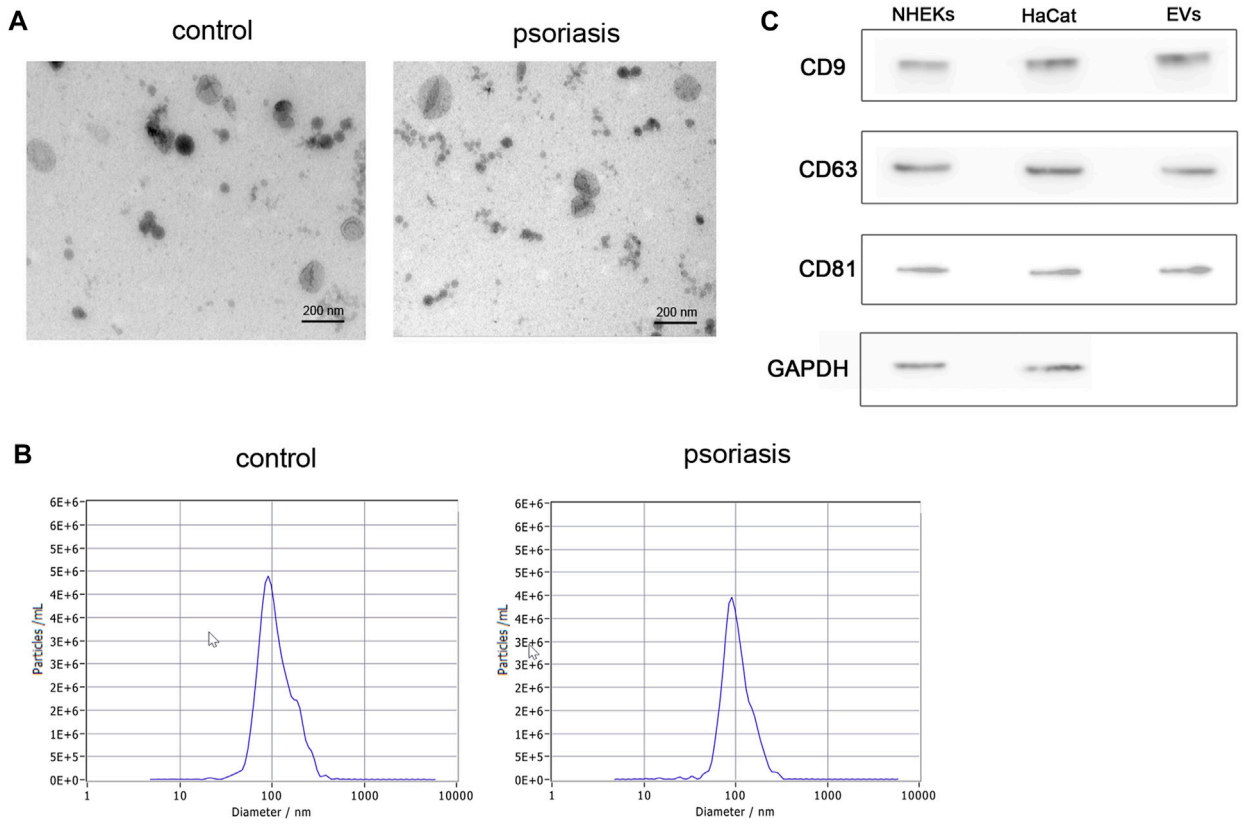


FIGURE 1
Isolation and characterization of serum EVs. **(A)** Serum EVs were observed using transmission electron microscopy (scale bar, 200 nm). Serum EVs exhibited a typical artificial cup shape appearance; **(B)** the diameter of serum EVs was detected by nanoparticle tracking analysis (n = 3); **(C)** surface marker proteins of serum EVs were analyzed by western blot (n = 3), NHEKs and HaCat cells were used as poistive controls. EVs: extracellular vesicles, scale bar = 200 nm.

patients (n = 15) and healthy controls (n = 10) for 12 h. Optical density (OD) values at 450 nm were measured using a Synergy H1 Microplate Reader (BioTek, United States).

Enzyme-linked immunosorbent assay assay

After 24 h co-cultured with serum EVs, the culture supernatants were collected. Then, the expression levels of CCL20 and IL-8 secreted from NHEK cells were measured using specific ELISA kits (Elabscience, China). Absorbance at 450 nm was measured using an ELISA plate reader (Bio-Rad).

Statistical analysis

All data are presented as mean ± standard deviation. All experiments were performed at least 3 times. The differences between 2 groups were analyzed using Student's t-test with SPSS 17_0 analytical software (IBM, Armonk, NY, United States) and GraphPad Prism 7.00 (GraphPad Software, La Jolla, CA, United States). The target genes in GO and pathway were using hypergeometric test for statistics and p-value was corrected by multiple hypothesis tests. p-values < 0.05 were considered significant.

Results

Isolation and characterization of serum extracellular vesicles

To investigate the characterization of serum EVs in psoriasis, we isolated serum EVs from psoriasis patients and healthy controls using differential ultracentrifugation. Using the methods of transmission electron microscopy (TEM) and nanoparticle tracking analysis (NTA), we confirmed that purified serum EVs obtained from psoriasis patients and healthy controls both exhibited a typical cup-shaped structure

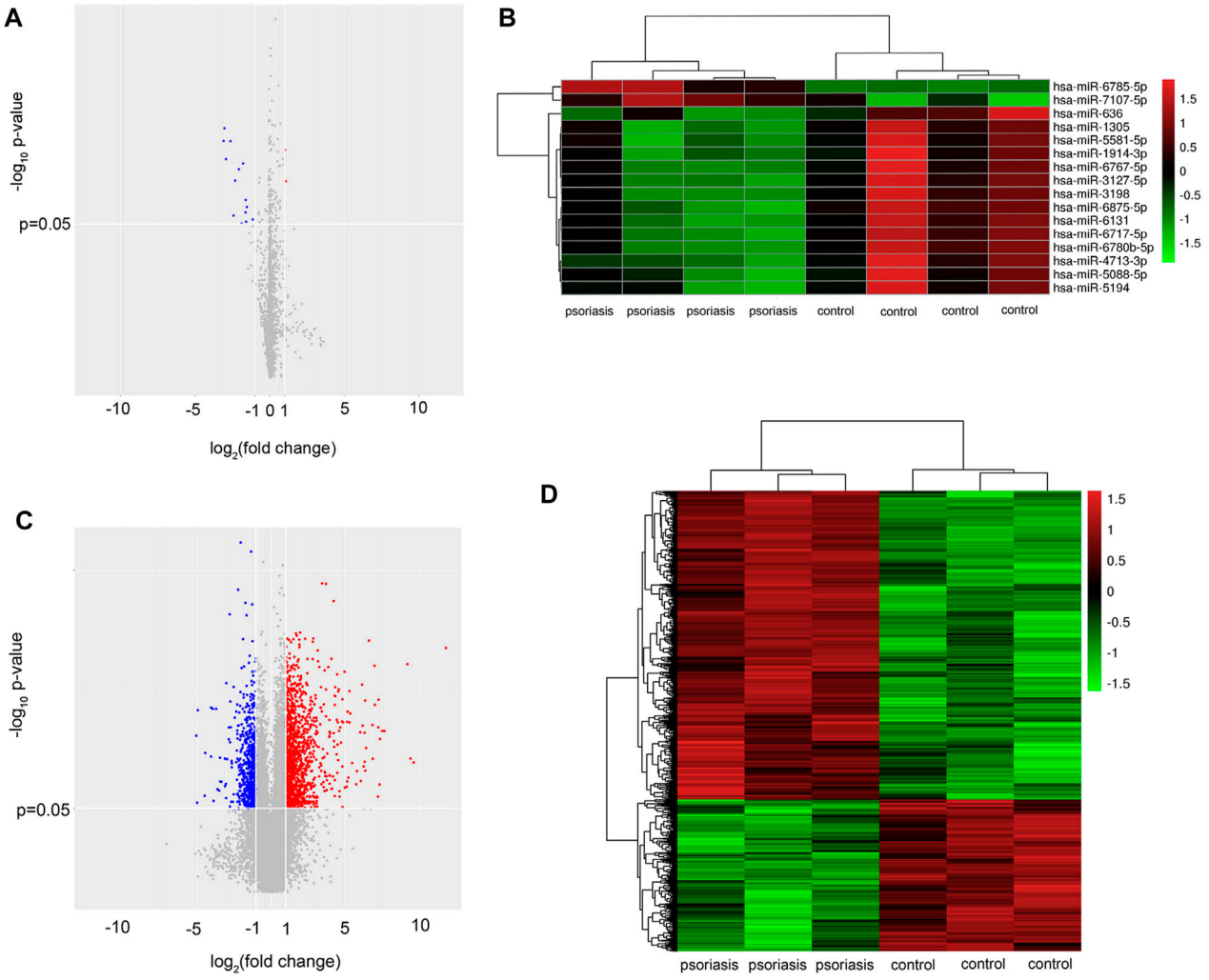


FIGURE 2
Serum EVs from psoriasis patients exhibit a specific miRNA profile. **(A)** Volcano plot of differentially expressed miRNAs from serum EVs between psoriasis and normal control (n = 4); **(B)** Hierarchical clustering analysis of the differentially expressed miRNAs (n = 4). Red and green colours indicated high and low expression, respectively; **(C)** volcano plot of differentially expressed mRNAs in skin lesions between psoriasis and normal control (n = 3). **(D)** Hierarchical clustering analysis of the differentially expressed mRNAs (n = 3). Red and green colours indicated high and low expression, respectively. miRNA: microRNA; EVs: extracellular vesicles.

(Figure 1A), and their average diameters were around 100 nm (Figure 1B). In addition, as shown in Figure 1C, the expression of CD9, CD63, and CD81 was identified in serum EVs by western blot analysis.

Serum extracellular vesicles from psoriasis patients exhibit a specific miRNA profile

To determine miRNA profiling of serum EVs in psoriasis, the expression of EVs miRNAs was detected using miRNA microarray analysis (n = 4). The results showed remarkable differences in the serum EVs miRNA profiles, with 16 differentially expressed miRNAs in serum EVs from psoriasis compared with healthy controls (GSE200637). Among 16 differentially expressed miRNAs, 2 were upregulated and 14 were down-regulated (fold change>2, $p < 0.05$, Figures 2A,B), (Supplementary Table S1).

To further investigate the potential role of serum EVs miRNAs in the activation of keratinocytes, we performed mRNA microarray analysis in psoriatic skin lesions and healthy controls (n = 3). According to the mRNA microarray results, a total of 1,725 dysregulated mRNAs were identified. Among 1,725 dysregulated mRNAs, 1,157 mRNAs were upregulated and 568 mRNAs were downregulated (fold change>2, $p < 0.05$, Figures 2C,D), (Supplementary Table S2).

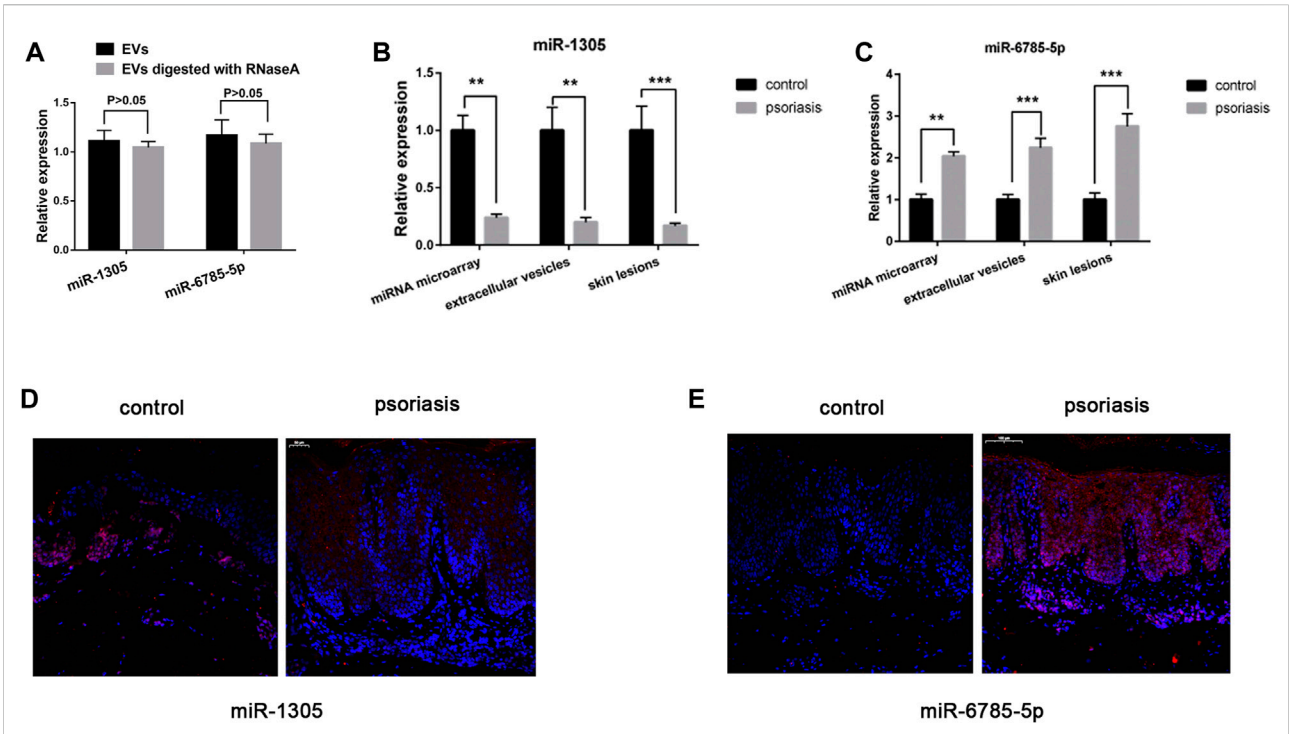


FIGURE 3
MiRNA microarray validation. **(A)** EVs digested with or without RNaseA, and miR-1305 and miR-6785-5p were detected by qRT-PCR. **(B)**The expression of miR-1305 was detected by miRNA microarray and validated in serum EVs and skin lesions respectively (n = 10). **(C)** The expression of miR-6785-5p was detected by miRNA microarray, and validated in serum EVs and skin lesions respectively (n = 10). **(D,E)** The expression of miR-1305 and miR-6785-5p in skin lesions were detected by FISH (n = 10). miRNA: microRNA; scale bar = 100 μm. EVs: extracellular vesicles; FISH: Fluorescence *in situ* hybridization analysis. **indicates $p < 0.01$ and ***indicates $p < 0.001$.

The mRNA microarray results suggested that there were a greater number of upregulated mRNAs in psoriatic skin lesions compared with healthy controls, which corresponded to the majority of miRNAs that were downregulated in serum EVs.

miRNA microarray validation

To validate the miRNA microarray results, we performed qRT-PCR (n = 10) and Fluorescence *in situ* hybridization (n = 10) to detect the expression of miR-1305 and miR-6785-5p which predicted closely related with inflammation responses. We digested EVs with RNaseA and reperformed qRT-PCR, the results indicated that miR-1305 and miR-6785-5p are present inside of EVs not on the surface of EVs (Figure 3A). The results of Figures 3B,C showed that the expression of miR-1305 was downregulated in both serum EVs and skin lesions (fold change = 0.2 in EVs, $p < 0.01$; fold change = 0.17 in skin lesions, $p < 0.001$) and miR-6785-5p was upregulated (fold change = 2.24 in EVs, $p < 0.001$; fold change = 2.76 in skin lesions, $p < 0.001$). As shown in Figures 3D,E, the results of fluorescence *in situ* hybridization (FISH) were consistent with qRT-PCR and showed that miR-1305 and miR-6785-5p were differentially expressed in the epidermis.

Potential functions of miR-1305 and miR-6785-5p

To explore the potential functions of differentially expressed miRNAs in serum EVs, we predicted the target genes of the validated miR-1305 and miR-6785-5p using bioinformatics analysis (TargetScanHuman 7.1 and miRDB-MicroRNA Target Prediction Database). As miR-1305 was downregulated and miR-6785-5p was upregulated in serum EVs and psoriatic skin lesions, the predicted target genes of miR-1305 and miR-6785-5p were intersected with upregulated and downregulated mRNAs which were detected by mRNA microarray (GSE181318), respectively. The results showed that 77 upregulated mRNAs might be regulated by miR-1305 and 83 downregulated mRNAs might be regulated by miR-6785-5p (Figures 4A,B); Supplementary Tables S3, S4). In addition, we mapped the miRNA-mRNA networks of miR-1305 and miR-6785-5p to study their regulatory mechanisms with the 77 upregulated mRNAs and 83 downregulated mRNAs, respectively (Figures 4C,D); Supplementary Tables S5,S6).

To clarify the potential biological functions of the 77 upregulated mRNAs and 83 downregulated mRNAs and the predicted involved signalling pathways, we annotated each gene of the intersection based on the Gene Ontology and KEGG

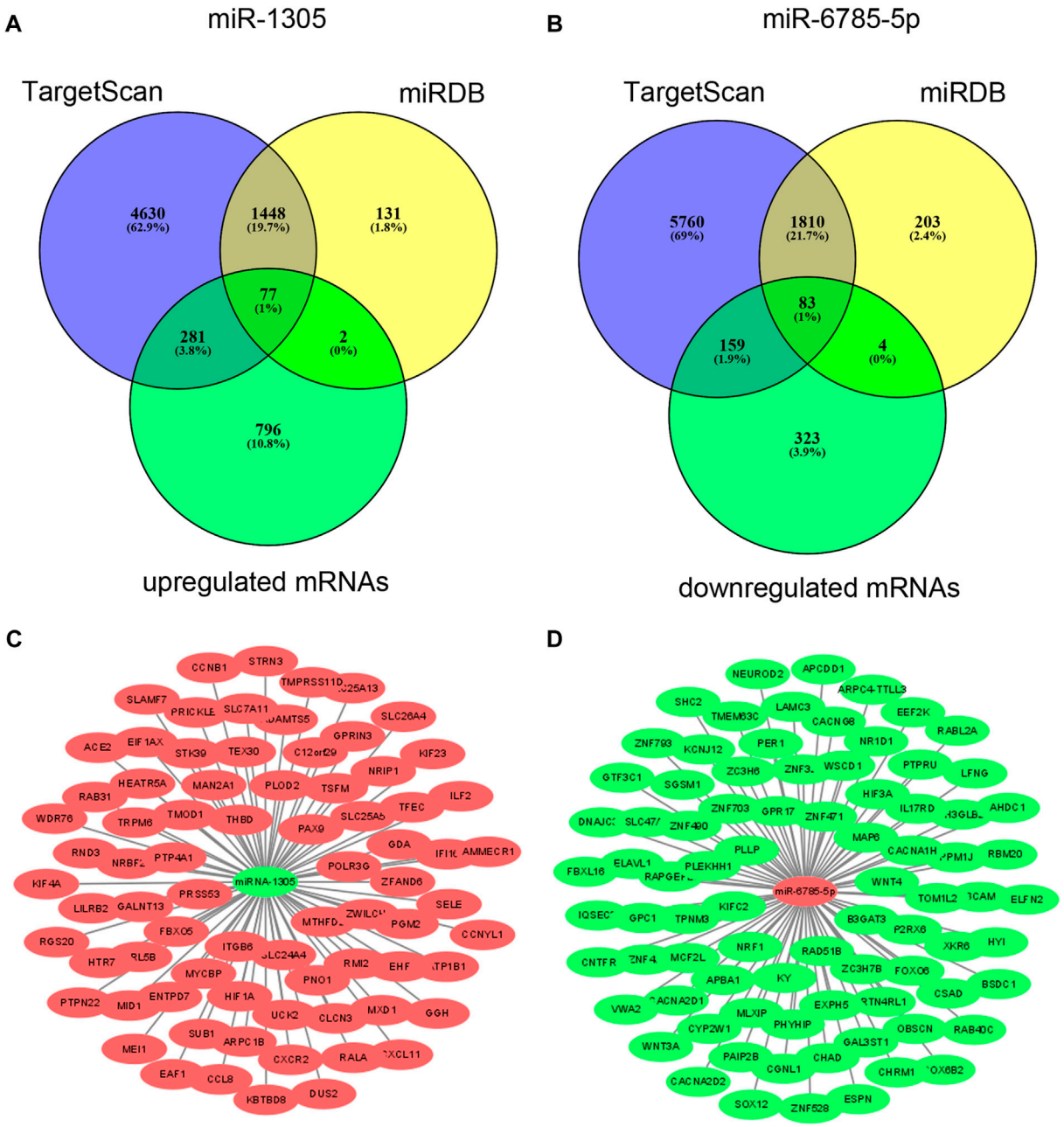


FIGURE 4
Predicted target genes of miR-1305 and miR-6785-5p, and miRNA-mRNA network analysis. **(A,B)** Target genes of miR-1305 and miR-6785-5p were predicted by TargetScan and miRDB, and then intersected with upregulated and downregulated mRNAs which detected by mRNA microarray respectively. **(C,D)** miRNA-mRNA network analysis of miR-1305 and miR-6785-5p, Red and green colours indicated high and low expression, respectively. miRNA: microRNA; EVs: extracellular vesicles. miRna-mRNA net-work analysis was mapped using Cytoscape 3.9.0 software.

database. Then, GO analysis and KEGG pathway analysis were performed to predict the function of miR-1305 and miR-6785-5p, respectively. The results showed that both miR-1305 and miR-6785-5p were involved in regulating the inflammation response and the signalling pathway related to psoriasis, including positive regulating of type 1 interferon production, autophagy, cytokine-cytokine receptor interaction, Th17 cell differentiation, IL-17 receptor activity, Wnt signalling pathway, MAPK signalling pathway and so on (Supplementary Figures S1A,D). These results indicated that miR-1305 and miR-6785-5p might be involved in the pathogenesis of psoriasis.

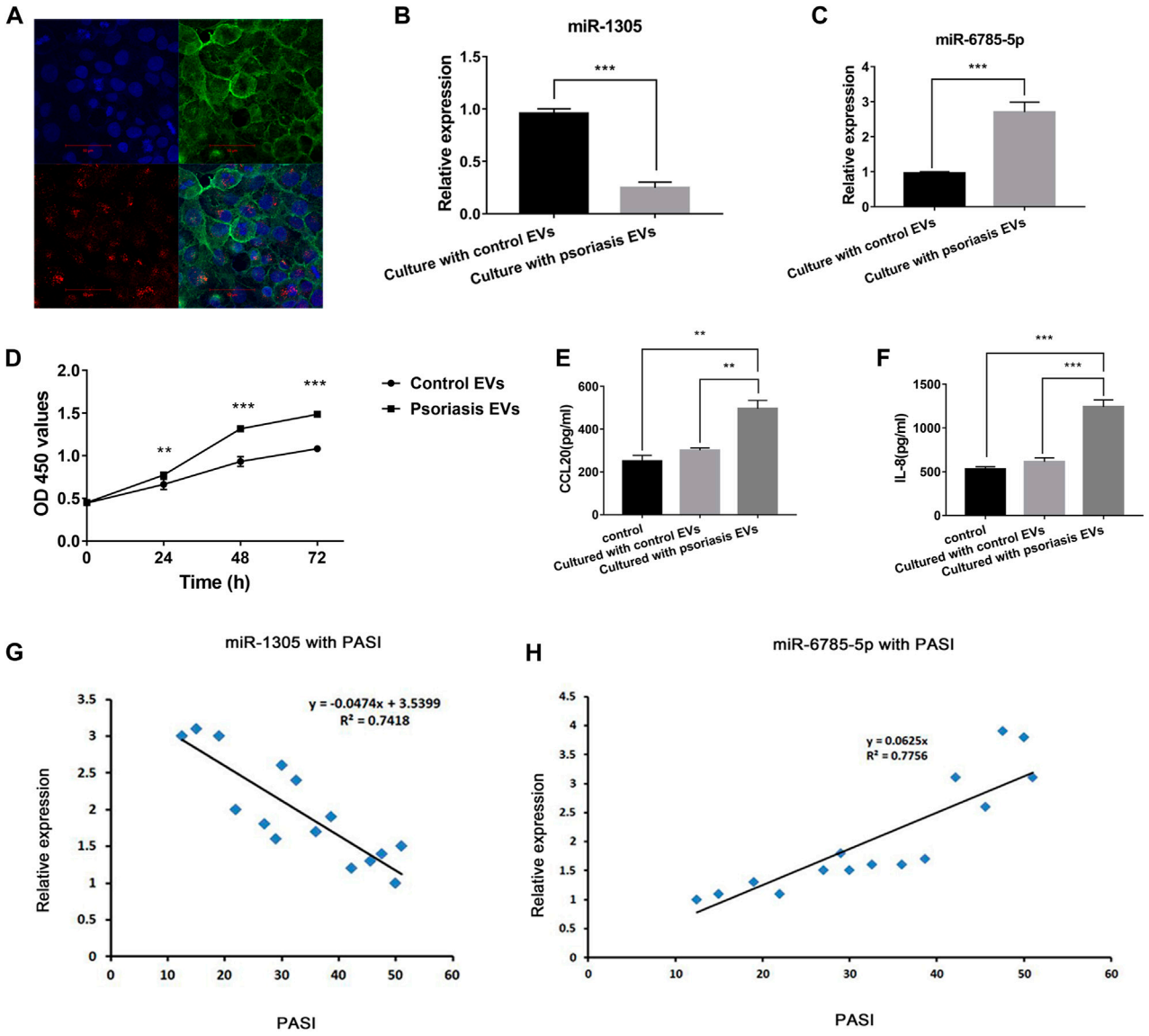


FIGURE 5
Serum EVs-keratinocytes interaction and serum EVs miR-1305 and miR-6785-5p might be biomarkers of psoriasis. **(A)** Uptake experiment of serum EVs; PKH26-labeled hucMSCs-Exo (red), cytoskeleton was labelled with phalloidin stain (green), and nuclei were stained with DAPI (blue). Scale bar = 50 μm; EVs: extracellular vesicles. **(B,C)** The expression of miR-1305 and miR6785-5p in NHEKs after co-cultured with serum EVs from psoriasis patients and controls. **(D)** The effects of serum EVs from psoriasis patients and controls on the proliferation of NHEKs. OD450: optical density measured at 450 nm. **(E,F)** The effects of serum EVs from psoriasis patients and controls on the secretion of CCL20 and IL-8 by NHEKs. **(G,H)** The expression of serum EVs miR-1305 and miR-6785-5p with PASI. miRNA: microRNA, PASI: plaques are graded based on three criteria: redness (R), thickness (T), and scaliness (S). Severity is rated for each index on a 0–4 scale (0 for no involvement; 4 for severe involvement). The body is divided into 4 regions: head (h), upper extremities (u), trunk (t), and lower extremities (l). In each of these areas, the fraction of total surface area affected is graded on a 0–6 scale (0, no involvement; up to 6 for >90% involvement). The various body regions are weighted to reflect their respective proportion of body surface area. The composite PASI score can then be calculated: PASI = 0.1(Rh + Th + Sh)Ah + 0.2(Ru + Tu + Su)Au + 0.3(Rt + Tt + St)At + 0.4(Rl + Tl + Sl)Al.

Serum extracellular vesicles promote keratinocytes proliferation and inflammation response

To investigate whether serum EVs were involved in the activation of keratinocytes, we performed a serum EVs-keratinocytes interaction experiment. Purified serum EVs were isolated and labelled with PKH26 Red Fluorescent. Next, the labelled EVs were co-cultured with NHEK cells for 12 h. Then, the cells were fixed with 4% formaldehyde and stained with phalloidin-iFluor 488 reagent and DAPI. Using a confocal microscope, we verified that serum EVs could be taken up by NHEK cells (Figure 5A).

Next, we detected the expression of miR-1305 and miR-6785-5p in NHEK cells after co-cultured with serum EVs obtained

from psoriasis patients and healthy controls. The results showed that miR-1305 was downregulated and miR-6785-5p was upregulated after co-cultured with serum EVs obtained from psoriasis compared with controls (Figures 5B,C). As the intrinsic cell targeting properties of EVs, our findings assumed that serum EVs might not only directly communicate with neighbouring cells, such as T cells or B cells, but also communicate with the activation of keratinocytes in psoriasis *via* the loaded miRNAs, including miR-1305 and miR-6785-5p.

Then, we explored the role of serum EVs in keratinocyte proliferation and inflammatory response. CCK-8 assay showed that serum EVs obtained from psoriasis significantly promoted the proliferation of keratinocytes compared with that of control groups (Figure 5D). Similarly, the results of ELISA analysis showed the secretion of CCL20 and IL-8 was significantly increased compared with that of control groups (Figures 5E,F). These findings suggested that serum EVs might be involved in the activation of keratinocytes in psoriasis.

Serum EVs miR-1305 and miR-6785-5p are related to the disease severity and might be biomarkers of psoriasis.

To further confirm the relationship of serum EVs miR-1305 and miR-6785-5p with the severity of psoriasis, we measured the correlation of miR-1305 and miR-6785-5p in serum EVs with PASI scores. As Figures 5G,H showed, the increase of miR-6785-5p was positively correlated with the PASI score. In contrast, the expression of miR-1305 in serum EVs was negatively correlated with disease severity. These results indicated that serum EVs miR-1305 and miR-6785-5p might serve as biomarkers of psoriasis.

Discussion

It is well known that a complex interplay among keratinocytes, immune cells and inflammatory mediators contributes to the pathogenesis of psoriasis (Raeber et al., 2018; Rendon and Schakel, 2019). In recent years, EVs have become a subject of intense study. EVs, such as microvesicles and exosomes, act as cell-to-cell communication vectors and potential biomarkers for diseases (Liu et al., 2019). The impacts of EVs transfer have been reported in many critical cellular processes including cell-to-cell communication and immune response regulation (Robbins et al., 2016). EVs are released by donor cells and can be taken up by recipient cells. EVs can carry RNAs and proteins which may affect the phenotype of the recipient cells (Mulcahy et al., 2014). Although the role of secreted cytokines and chemokines has been well documented in psoriasis, the potential role of EVs has not been sufficiently investigated in this disease (Deng et al., 2016; Yan et al., 2019). The communication between serum EVs and keratinocytes remains less known. Here, we found keratinocytes could take up serum EVs. In addition, considering the intrinsic cell targeting properties of EVs, we assumed that serum EVs might be taken up by keratinocytes and serum EVs loaded miRNAs might be involved in regulating the mRNAs expression and activation of keratinocytes in psoriasis.

EVs are cell-secreted lipid bilayer membranous particles with heterogeneous size and composition (Li et al., 2020). They derive either from the endosomal compartment (exosomes) or as a result of shedding from the plasma membrane (microvesicles, endosomes and apoptotic bodies) (O'Brien et al., 2020). EVs may collapse during drying, resulting in a cup-shaped morphology (Raposo and Stoorvogel, 2013). The average diameter of EVs ranges from nanometres to a few micrometres, and the variety of their biological content includes lipids, proteins, and RNAs (Fabbiano et al., 2020). The major proteins involved with EVs biogenesis include CD63, CD81, CD9, etc (Veziroglu and Mias, 2020). Here in our study, using the methods of TEM and NTA, we verified that purified serum EVs exhibited a typical cup-shaped structure and the mean diameters were around 100 nm. In addition, the protein markers CD9, CD63, and CD81 were detected in serum EVs and the results were consistent with previous studies.

MiRNAs are important gene regulatory molecules and regulate vital cellular processes and inflammation regulation (Fabian et al., 2010). They modulate protein expression by inhibiting mRNA translation and stability. Our previous studies have verified that differentially expressed miRNAs in skin lesions were related to psoriasis, including miR-145-5p, miR-20a-3p, and miR-548a-3p (Li et al., 2018; Zhao et al., 2018; Yan et al., 2019). Recently, EVs miRNAs are considered to be involved in the pathogenesis of psoriasis. A previous study has identified that small EVs containing miR-381-3p from keratinocytes promotes Th1/Th17 polarization in psoriasis (Jiang et al., 2021). Another report identified circulating miRNAs in EVs as potential biomarkers for psoriatic arthritis in patients with psoriasis (Pasquali et al., 2020). However, the role of serum EVs in the pathogenesis of psoriasis remains less known.

Here, we found that miRNAs are indeed abnormally expressed in serum EVs in psoriasis. In addition, we confirmed that keratinocytes could take up the serum EVs. The result laid the foundation for studying the role of serum EVs in regulating the activation of keratinocytes and skin inflammation. As the intrinsic cell targeting properties of EVs, the results of miRNA microarray results and the EVs-keratinocytes interaction experiment indicated that there might be interactions between the dysregulated miRNAs derived from EVs and mRNAs in skin lesions. Furthermore, through the bioinformatics analysis, we found the dysregulated miRNAs derived from EVs might be involved in the activation of keratinocytes in psoriasis.

To further study the role of differential expression miRNAs derived from EVs in psoriasis, we found that miR-1305 and miR-6785-5p were closely related to inflammation response. Previous studies have reported that miR-1305 and miR-6785-5p play

important regulating roles in several diseases (Yu et al., 2019; Liu et al., 2020; Su et al., 2020). However, their functions in psoriasis have not been studied before. In this study, we found that miR-1305 and miR-6785-5p were participating in regulating inflammation response associated with psoriasis. Furthermore, the results of the bioinformatic analysis suggested that miR-1305 and miR-6785-5p might be involved in regulating the activation and differentiation of Th17 and Th1, which were not reported before. In addition, our findings indicated that miR-1305 and miR-6785-5pmight play an important role in the pathogenesis of psoriasis (Chang et al., 2018; Roy et al., 2018; Wu et al., 2018).

It is generally accepted that EVs can be taken up and impact the functions of recipient cells. miRNAs in extracellular vesicles (EVs), may mediate paracrine and endocrine communication between different tissues and thus modulate gene expression and the function of distal cells (Mori et al., 2019). Here, we confirmed serum EVs regulated the expression of miR-1305 and miR-6785-5p in NHEKs and promoted NHEKs proliferation and proinflammatory cytokines secretion, including CCL20 and IL-8. CCL20 plays an important role in psoriasis by recruiting CCR6+Th17 cells into the lesional skin (Furue et al., 2020). IL-8 is also an important factor in psoriasis, which is characterized by proliferation of keratinocytes, neutrophil infiltration and angiogenesis (Wu et al., 2017). These results indicated that psoriasis serum EVs were involved in the activation of keratinocytes and skin inflammation in psoriasis. Furthermore, the functions of serum EVs may be through loading and transferring miR-1305 and miR-6785-5p.

Increasing studies have focused on extracellular miRNAs as potential biomarkers, since they are stable and can be detected in the blood, urine, or other body fluids (Grasedieck et al., 2012; Mori et al., 2019). In this study, by analyzing the correlation between the expression of serum EVs miR-1305 and miR-6785-5p with the Psoriasis Area and Severity Index (PASI) score, we found the expression of serum EVs miR-6785-5p was positively correlated with the PASI score. In contrast, the expression of serum EVs miR-1305 was negatively correlated with the PASI score. These results indicated that serum EVs miR-1305 and miR-6785-5p can be served as biomarkers in psoriasis. In addition, miR-1305 and miR-6785-5p may be potential candidates for therapy of psoriasis in the future.

However, this study also has some limitations. Firstly, the tissue sample sizes were relatively small, and larger samples are required in the future to verify our findings. Secondly, the biomarkers only focused on psoriasis. The specificity of the biomarkers miR-1305 and miR-6785-5p in other inflammatory diseases should be investigated. Thirdly, the molecular mechanism of miR-1305 and miR-6785-5p in psoriasis has not yet been fully revealed in this study, and the remaining 14 miRNAs were left out in this study, which limited the investigation of their functions in this study. Finally, due to limitation of the sample size, we only showed the correlation of miR-1305 and miR-6785-5p in serum EVs with PASI scores.

Conclusion

In conclusion, the current study may suggest that serum EVs might be involved in the activation of keratinocytes through loaded miRNAs in psoriasis. Serum EVs miR-1305 and miR-6785-5p may be associated with psoriasis. This study may provide insight for further into the development and treatment of psoriasis.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: GSE200637 and GSE181318.

Ethics statement

The studies involving human participants were reviewed and approved by the ethics committee of Shandong University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

JY contributed to the conception, design of the work, analysis, and interpretation of data, draft this manuscript. YZ and RW contributed to the design of the work, analysis, and interpretation of data. XL, SH, HZ, and HW contributed to the acquisition and interpretation of data. QS contributed to the conception, design of the work. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (Grant Nos. 81972937 and 82003344) and the Natural Science Foundation of Shandong Province (ZR2020QH137).

Acknowledgments

We would like to acknowledge the helpful comments on this article received from our reviewers. We thank the staff of the laboratory of Basic Medical Sciences for their help during the experiment.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.891465/full#supplementary-material

References

Boehncke, W. H., and Schon, M. P. (2015). Psoriasis. *Lancet* 386 (9997), 983–994. doi:10.1016/S0140-6736(14)61909-7

Chang, H. W., Yan, D., Singh, R., Liu, J., Lu, X., Ucmak, D., et al. (2018). Alteration of the cutaneous microbiome in psoriasis and potential role in Th17 polarization. *Microbiome* 6 (1), 154. doi:10.1186/s40168-018-0533-1

Chen, H. L., Lo, C. H., Huang, C. C., Lu, M. P., Hu, P. Y., Chen, C. S., et al. (2021). Galectin-7 downregulation in lesional keratinocytes contributes to enhanced IL-17A signaling and skin pathology in psoriasis. *J. Clin. Invest.* 131 (1), 130740. doi:10.1172/JCI130740

Cheung, K. L., Jarrett, R., Subramaniam, S., Salimi, M., Gutowska-Owsiak, D., Chen, Y. L., et al. (2016). Psoriatic T cells recognize neolipid antigens generated by mast cell phospholipase delivered by exosomes and presented by CD1a. *J. Exp. Med.* 213 (11), 2399–2412. doi:10.1084/jem.20160258

Deng, Y., Chang, C., and Lu, Q. (2016). The inflammatory response in psoriasis: A comprehensive review. *Clin. Rev. Allergy Immunol.* 50 (3), 377–389. doi:10.1007/s12016-016-8535-x

Fabbiano, F., Corsi, J., Gurrieri, E., Trevisan, C., Notarangelo, M., and D'Agostino, V. G. (2020). RNA packaging into extracellular vesicles: An orchestra of RNA-binding proteins? *J. Extracell. Vesicles* 10 (2), e12043. doi:10.1002/jev2.12043

Fabian, M. R., Sonenberg, N., and Filipowicz, W. (2010). Regulation of mRNA translation and stability by microRNAs. *Annu. Rev. Biochem.* 79, 351–379. doi:10.1146/annurev-biochem-060308-103103

Frischknecht, L., Vecellio, M., and Selmi, C. (2019). The role of epigenetics and immunological imbalance in the etiopathogenesis of psoriasis and psoriatic arthritis. *Ther. Adv. Musculoskelet. Dis.* 11, 1759720X19886505. doi:10.1177/1759720X19886505

Furue, K., Ito, T., Tsuji, G., Nakahara, T., and Furue, M. (2020). The CCL20 and CCR6 axis in psoriasis. *Scand. J. Immunol.* 91 (3), e12846. doi:10.1111/sji.12846

Grasedieck, S., Scholer, N., Bommer, M., Niess, J. H., Tumani, H., Rouhi, A., et al. (2012). Impact of serum storage conditions on microRNA stability. *Leukemia* 26 (11), 2414–2416. doi:10.1038/leu.2012.106

Griffiths, C. E. M., Armstrong, A. W., Gudjonsson, J. E., and Barker, J. (2021). Psoriasis. *Lancet* 397 (10281), 1301–1315. doi:10.1016/S0140-6736(20)32549-6

Guinea-Viniegra, J., Jimenez, M., Schonthaler, H. B., Navarro, R., Delgado, Y., Concha-Garzon, M. J., et al. (2014). Targeting miR-21 to treat psoriasis. *Sci. Transl. Med.* 6 (225), 225re1. doi:10.1126/scitranslmed.3008089

Jiang, M., Fang, H., Dang, E., Zhang, J., Qiao, P., Yu, C., et al. (2021). Small extracellular vesicles containing miR-381-3p from keratinocytes promote T helper type 1 and T helper type 17 polarization in psoriasis. *J. Invest. Dermatol.* 141 (3), 563–574. doi:10.1016/j.jid.2020.07.009

Jiang, M., Ma, W. Y., Gao, Y. M., Jia, K., Zhang, Y., Liu, H. D., et al. (2017). IL-22-induced miR-122-5p promotes keratinocyte proliferation by targeting Sprouty2. *Exp. Dermatol.* 26 (4), 368–374. doi:10.1111/exd.13270

Kim, D., Kim, H. J., Baek, J. O., Roh, J. Y., and Jun, H. S. (2021). Lysophosphatidic acid mediates imiquimod-induced psoriasis-like symptoms by promoting keratinocyte proliferation through LPAR1/ROCK2/PI3K/AKT signaling pathway. *Int. J. Mol. Sci.* 22 (19), 10777. doi:10.3390/ijms221910777

Kim, K. M., Abdelmohsen, K., Mustapic, M., Kapogiannis, D., and Gorospe, M. (2017). RNA in extracellular vesicles. *WIREs. RNA* 8 (4), e1413. doi:10.1002/wrna.1413

Li, R., Qiao, M., Zhao, X., Yan, J., Wang, X., and Sun, Q. (2018). MiR-20a-3p regulates TGF-β1/Survivin pathway to affect keratinocytes proliferation and apoptosis by targeting SFMBT1 *in vitro*. . *Signal.* 49, 95–104. doi:10.1016/j.cellsig.2018.06.003

Li, Y. J., Wu, J. Y., Wang, J. M., Hu, X. B., and Xiang, D. X. (2020). Emerging strategies for labeling and tracking of extracellular vesicles. *J. Control. Release* 328, 141–159. doi:10.1016/j.jconrel.2020.08.056

Liu, T., Zhang, Q., Zhang, J., Li, C., Miao, Y. R., Lei, Q., et al. (2019). EVmiRNA: A database of miRNA profiling in extracellular vesicles. *Nucleic Acids Res.* 47 (D1), D89–D93. doi:10.1093/nar/gky985

Liu, W., Zhuang, R., Feng, S., Bai, X., Jia, Z., Kapora, E., et al. (2020). Long non-coding RNA ASB16-AS1 enhances cell proliferation, migration and invasion via functioning as a ceRNA through miR-1305/Wnt/β-catenin axis in cervical cancer. *Biomed. Pharmacother.* 125, 109965. doi:10.1016/j.biopha.2020.109965

Lu, T. X., and Rothenberg, M. E. (2018). MicroRNA. *J. Allergy Clin. Immunol.* 141 (4), 1202–1207. doi:10.1016/j.jaci.2017.08.034

Mori, M. A., Ludwig, R. G., Garcia-Martin, R., Brandao, B. B., and Kahn, C. R. (2019). Extracellular miRNAs: From biomarkers to mediators of physiology and disease. *Cell. Metab.* 30 (4), 656–673. doi:10.1016/j.cmet.2019.07.011

Mulcahy, L. A., Pink, R. C., and Carter, D. R. (2014). Routes and mechanisms of extracellular vesicle uptake. *J. Extracell. Vesicles* 3, 24641. doi:10.3402/jev.v3.24641

Ni, X., and Lai, Y. (2020). Keratinocyte: A trigger or an executor of psoriasis? *J. Leukoc. Biol.* 108 (2), 485–491. doi:10.1002/JLB.5MR0120-439R

Nussbaum, L., Chen, Y. L., and Ogg, G. S. (2021). Role of regulatory T cells in psoriasis pathogenesis and treatment. *Br. J. Dermatol.* 184 (1), 14–24. doi:10.1111/bjd.19380

O'Brien, K., Breyne, K., Ughetto, S., Laurent, L. C., and Breakefield, X. O. (2020). RNA delivery by extracellular vesicles in mammalian cells and its applications. *Nat. Rev. Mol. Cell. Biol.* 21 (10), 585–606. doi:10.1038/s41580-020-0251-y

Papayannakos, C. J., DeVoti, J. A., Israr, M., Alsudani, H., Bonagura, V., and Steinberg, B. M. (2021). Extracellular vesicles produced by primary human keratinocytes in response to TLR agonists induce stimulus-specific responses in antigen-presenting cells. *Cell. Signal.* 83, 109994. doi:10.1016/j.cellsig.2021.109994

Pasquali, L., Svedbom, A., Srivastava, A., Rosen, E., Lindqvist, U., Stahle, M., et al. (2020). Circulating microRNAs in extracellular vesicles as potential biomarkers for psoriatic arthritis in patients with psoriasis. *J. Eur. Acad. Dermatol. Venereol.* 34 (6), 1248–1256. doi:10.1111/jdv.16203

Raeber, M. E., Zurbuchen, Y., Impellizzieri, D., and Boyman, O. (2018). The role of cytokines in T-cell memory in health and disease. *Immunol. Rev.* 283 (1), 176–193. doi:10.1111/imr.12644

Raposo, G., and Stoorvogel, W. (2013). Extracellular vesicles: Exosomes, microvesicles, and friends. *J. Cell. Biol.* 200 (4), 373–383. doi:10.1083/jcb.201211138

Rendon, A., and Schakel, K. (2019). Psoriasis pathogenesis and treatment. *Int. J. Mol. Sci.* 20 (6), E1475. doi:10.3390/ijms20061475

Robbins, P. D., Dorronsoro, A., and Booker, C. N. (2016). Regulation of chronic inflammatory and immune processes by extracellular vesicles. *J. Clin.* . 126 (4), 1173–1180. doi:10.1172/JCI81131

Roy, A. K., Roy, P. K., and Grigorieva, E. (2018). Mathematical insights on psoriasis regulation: Role of Th_1 and Th_2 cells. *Math. Biosci. Eng.* 15 (3), 717–738. doi:10.3934/mbe.2018032

Shao, S., Fang, H., Li, Q., and Wang, G. (2020). Extracellular vesicles in inflammatory skin disorders: From pathophysiology to treatment. *Theranostics* 10 (22), 9937–9955. doi:10.7150/thno.45488

Shao, S., Fang, H., Zhang, J., Jiang, M., Xue, K., Ma, J., et al. (2019). Neutrophil exosomes enhance the skin autoinflammation in generalized pustular psoriasis via activating keratinocytes. *FASEB J.* 33 (6), 6813–6828. doi:10.1096/fj.201802090RR

Srivastava, A., Nikamo, P., Lohcharoenkal, W., Li, D., Meisgen, F., Xu Landen, N., et al. (2017). MicroRNA-146a suppresses IL-17-mediated skin inflammation and is genetically associated with psoriasis. *J. Allergy Clin. Immunol.* 139 (2), 550–561. doi:10.1016/j.jaci.2016.07.025

Su, Y., Feng, W., Shi, J., Chen, L., Huang, J., and Lin, T. (2020). circRIP2 accelerates bladder cancer progression via miR-1305/Tgf-β2/smad3 pathway. *Mol. Cancer* 19 (1), 23. doi:10.1186/s12943-019-1129-5

Than, U. T. T., Leavesley, D. I., and Parker, T. J. (2019). Characteristics and roles of extracellular vesicles released by epidermal keratinocytes. *J. Eur. Acad. Dermatol. Venereol.* 33 (12), 2264–2272. doi:10.1111/jdv.15859

Tsai, Y. C., and Tsai, T. F. (2017). Anti-interleukin and interleukin therapies for psoriasis: Current evidence and clinical usefulness. *Ther. Adv. Musculoskelet. Dis.* 9 (11), 277–294. doi:10.1177/1759720X17735756

Vader, P., Mol, E. A., Pasterkamp, G., and Schiffelers, R. M. (2016). Extracellular vesicles for drug delivery. *Adv. Drug Deliv. Rev.* 106, 148–156. doi:10.1016/j.addr.2016.02.006

Veziroglu, E. M., and Mias, G. I. (2020). Characterizing extracellular vesicles and their diverse RNA contents. *Front. Genet.* 11, 700. doi:10.3389/fgene.2020.00700

Wu, P., Ma, G., Zhu, X., Gu, T., Zhang, J., Sun, Y., et al. (2017). Cyr61/CCN1 is involved in the pathogenesis of psoriasis vulgaris via promoting IL-8 production by keratinocytes in a JNK/NF-κB pathway. *Clin. Immunol.* 174, 53–62. doi:10.1016/j.clim.2016.11.003

Wu, R., Zeng, J., Yuan, J., Deng, X., Huang, Y., Chen, L., et al. (2018). MicroRNA-210 overexpression promotes psoriasis-like inflammation by inducing Th1 and Th17 cell differentiation. *J. Clin.* . 128 (6), 2551–2568. doi:10.1172/JCI97426

Yan, J. J., Qiao, M., Li, R. H., Zhao, X. T., Wang, X. Y., and Sun, Q. (2019). Downregulation of miR-145-5p contributes to hyperproliferation of keratinocytes and skin inflammation in psoriasis. *Br. J. Dermatol.* 180 (2), 365–372. doi:10.1111/bjd.17256

Yu, C., Chen, D. Q., Liu, H. X., Li, W. B., Lu, J. W., and Feng, J. F. (2019). Rosmarinic acid reduces the resistance of gastric carcinoma cells to 5-fluorouracil by downregulating FOXO4-targeting miR-6785-5p. *Biomed. Pharmacother. = Biomedecine Pharmacother.* 109, 2327–2334. doi:10.1016/j.biopha.2018.10.061

Zhao, X., Li, R., Qiao, M., Yan, J., and Sun, Q. (2017). MiR-548a-3p promotes keratinocyte proliferation targeting PPP3R1 after being induced by IL-22. *Inflammation* 41 (2), 496–504. doi:10.1007/s10753-017-0705-3

Zhao, X., Li, R., Qiao, M., Yan, J., and Sun, Q. (2018). MiR-548a-3p promotes keratinocyte proliferation targeting PPP3R1 after being induced by IL-22. *Inflammation* 41 (2), 496–504. doi:10.1007/s10753-017-0705-3

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 15 August 2022
DOI 10.3389/fgene.2022.941221

Check for updates

OPEN ACCESS

EDITED BY
Yonghu Sun,
Shandong Provincial Hospital of Dermatology, China

REVIEWED BY
Adesh Kumar Saini,
Maharishi Markandeshwar University, India
Qixiang Shao,
Jiangsu University, China
Pattarin Tangtanatakul,
Chulalongkorn University, Thailand

*CORRESPONDENCE
Lingyun Sun,
lingyunsun@nju.edu.cn
Shuai Ding,
dingshuai@njglyy.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

RECEIVED 11 May 2022
ACCEPTED 11 July 2022
PUBLISHED 15 August 2022

CITATION
Li Z, Wang Z, Sun T, Liu S, Ding S and Sun L (2022), Identifying key genes in CD4⁺ T cells of systemic lupus erythematosus by integrated bioinformatics analysis.
Front. Genet. 13:941221.
doi: 10.3389/fgene.2022.941221

COPYRIGHT
© 2022 Li, Wang, Sun, Liu, Ding and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Identifying key genes in CD4⁺ T cells of systemic lupus erythematosus by integrated bioinformatics analysis

Zutong Li[1†], Zhilong Wang[2†], Tian Sun[1], Shanshan Liu[1], Shuai Ding[1*] and Lingyun Sun[1*]

[1]Department of Rheumatology and Immunology, The Affiliated Drum Tower Hospital of Nanjing University Medical School, Nanjing, China, [2]Department of Reproductive Medicine Center, The Affiliated Drum Tower Hospital of Nanjing University Medical School, Nanjing, China

Systemic lupus erythematosus (SLE) is a chronic autoimmune disease characterized by excessive activation of T and B lymphocytes and breakdown of immune tolerance to autoantigens. Despite several mechanisms including the genetic alterations and inflammatory responses have been reported, the overall signature genes in CD4⁺ T cells and how they affect the pathological process of SLE remain to be elucidated. This study aimed to identify the crucial genes, potential biological processes and pathways underlying SLE pathogenesis by integrated bioinformatics. The gene expression profiles of isolated peripheral CD4⁺ T cells from SLE patients with different disease activity and healthy controls (GSE97263) were analyzed, and 14 co-expression modules were identified using weighted gene co-expression network analysis (WGCNA). Some of these modules showed significantly positive or negative correlations with SLE disease activity, and primarily enriched in the regulation of type I interferon and immune responses. Next, combining time course sequencing (TCseq) with differentially expressed gene (DEG) analysis, crucial genes in lupus CD4⁺ T cells were revealed, including some interferon signature genes (ISGs). Among these genes, we identified 4 upregulated genes (*PLSCR1, IFI35, BATF2* and *CLDN5*) and 2 downregulated genes (*GDF7* and *DERL3*) as newfound key genes. The elevated genes showed close relationship with the SLE disease activity. In general, our study identified 6 novel biomarkers in CD4⁺ T cells that might contribute to the diagnosis and treatment of SLE.

KEYWORDS

systemic lupus erythematosus, CD4⁺ T cells, disease activity, weighted gene co-expression network analysis, biomarker, type I interferon, immune response

Abbreviations: BATF, basic leucine zipper transcription factor; BBB, brain-blood-barrier; DEG, differentially expressed gene; ER, endoplasmic reticulum; GO, Gene Ontology; HC, healthy controls; IFN, interferon; ISG, interferon signature gene; KEGG: Kyoto Encyclopedia of Genes and Genomes; LN: lupus nephritis; ME, module eigengene; NMI, N-myc-interactor; PBMC, peripheral blood mononuclear cell; PCA, principal component analysis; SLE, systemic lupus erythematosus; SLEDAI, systemic lupus erythematosus disease activity index; TCseq, time course sequencing analysis; TOM, topological overlap matrix; WGCNA, weighted gene co-expression network analysis.

Introduction

Systemic lupus erythematosus (SLE) is a chronic autoimmune disease which affects diffuse connective tissues and organs, including skin, joints and kidneys. SLE is characterized by excessive activation of T and B lymphocytes and breach of immune tolerance to autoantigens, which trigger the production of autoantibodies and lead to the immune-complex related inflammation in multiple organs and tissues (Tsokos, 2011). Genetic factors and environmental triggers are believed to play important roles in the pathogenesis and progression of SLE (Goulielmos et al., 2018; Dorner and Furie, 2019). Nevertheless, the pathogenic mechanisms of SLE have not been fully understood yet. Thus, the time-course of disease flares, remission and progression is unpredictable (Obermoser and Pascual, 2010). Besides, SLE is also a highly heterogeneous disease in terms of diverse clinical manifestations and severity, which presents a challenge to the clinicians and researchers. It would therefore be of great value to explore the molecular signatures underlying different clinical phenotypes, as it could aid in accurate diagnosis, disease activity assessment and clinical management of SLE.

The mutation and abnormal expression of many vital genes also confer a predisposition to SLE, indicating the value of diagnosis or prognosis (Teruel and Alarcon-Riquelme, 2016; Luo et al., 2020). However, peripheral blood mononuclear cells (PBMCs) consist of a mixture of lymphocytes and monocytes, and rarely show a good enough discrepancy on transcriptomic profiles. Among the major peripheral immune cells in lupus, the autoreactive and pro-inflammatory $CD4^+$ T cells stimulate the differentiation, proliferation and maturation of B cells to enhance the production of autoantibodies, playing a key role in the pathogenesis and progression of SLE (Moulton and Tsokos, 2011; Zhao et al., 2018; Jang et al., 2021). The alterations in the signaling physiology and gene transcription lead to abnormalities in the phenotype of these cells (Mak and Kow, 2014; Yuan et al., 2022). Most of all, the peripheral blood cells of lupus patients demonstrated overexpression of the gene profiles induced by type I interferon (IFN), also known as interferon signature genes (ISGs) (Feng et al., 2015; Ronnblom, 2016; Postal et al., 2020). However, how these molecular signatures correlate with SLE activity awaits further characterization. The transcriptomic or translational profiles of lupus $CD4^+$ T cells can lead to a better understanding of pathogenic mechanisms of SLE, and aid in potential therapeutic targets identification in an unbiased manner.

Weighted gene co-expression network analysis (WGCNA) is a well-known method of systems biology for exploring and identifying the potential functional pathways and biomarkers for diagnosis and prognosis of complex diseases at the level of the genome (Langfelder and Horvath, 2008). This powerful bioinformatic tool has been widely used in various diseases, including SLE and other autoimmune diseases (Yan et al., 2018; Sun et al., 2019). Using WGCNA, Liu *et al.* found overexpressed small RNAs encoded by human endogenous retrovirus K in PBMCs that might be involved in the immune regulation and progression of SLE (Liu et al., 2021). Similarly, *IFI27* may be closely related to pathogenesis of SLE (Zhao et al., 2021). In lupus nephritis (LN), the potential gene expression biomarkers for diagnosis and prognosis were developed by integrating multiple differentially expressed gene (DEG) identification methods (Yao et al., 2020; Chen et al., 2021; Shen et al., 2021). In this study, through integrated bioinformatics analysis of high-throughput sequencing data, we well-characterized the gene expression profiles in $CD4^+$ T cells obtained from healthy controls (HC) and lupus patients, and identified crucial genes correlated with the severity of SLE. These findings could improve our understanding of the disease pathogenesis and provide new insights in identification of potential diagnostic and therapeutic molecular targets of SLE.

Materials and methods

Data acquisition and processing

We obtained the gene expression dataset GSE97263 with corresponding clinical information from the GEO database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE97263), which was performed with the platform of Illumina HiSeq 2500. This dataset contained isolated blood $CD4^+$ T cells from 14 HC, 14 active and 16 inactive SLE patients (Buang et al., 2021). The Systemic Lupus Erythematosus Disease Activity Index (SLEDAI score) was used for clinical classification. Inactive SLE was defined as a SLEDAI <4 and active SLEDAI >6. Processing and analysis of these collected data were conducted with the R software. The Ensemble IDs were converted into gene symbols with the bitr function in clusterProfiler (Wu et al., 2021), and the genes with average raw reads value less than 1 were removed. After data processing, 16,623 genes were matched, and all these genes were used for the following WGCNA after normalized with log transformed (in detail, log(edgeR:cpm(counts+1)).

WGCNA

A sample clustering tree map was first constructed to detect and eliminate outliers. Then, the "WGCNA" R package (Langfelder and Horvath, 2008) was used to construct the gene network with the dataset GSE97263. In detail, scale independence and mean connectivity were identified *via* the soft threshold power (β value) setting of 1–20. Meanwhile, soft threshold power was selected as the degree of scale

independence reached 0.85. Based on the selected soft threshold, the adjacency matrix was converted to topological overlap matrix (TOM) to construct the network. Then, we performed module identification using cutreeDynamic function with minClusterSize = 100, and the gene dendrogram and module color were established using the degree of TOM-based dissimilarity(1-TOM). Next, MergeCutHeight function was used for cutting the dendrogram in the process of module merging with the a 0.25 MEDissThres value, and 14 modules were finally harvested.

Identification of clinically significant modules

The Pearson correlation coefficient between the module eigengene (ME) and sample traits was calculated to find out the highly relevant module (hub module) associated with the development of SLE. Modules with top 2 corresponding correlation with $p < 0.05$ were identified. Then the correlationship between Module Membership (MM, the correlation of the module eigengene and the gene expression profile) and Gene Significance (GS, the correlation between the gene and the clinical phenotypes) in these modules was calculated and visualized with the "WGCNA" R package (Langfelder and Horvath, 2008). The significantly correlation GS and MM implied that hub genes of the modules tend to be highly correlated with clinical phenotypes.

Functional enrichment analysis

In order to identify the function of the selected genes in the pathogenesis and development of SLE, Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment analysis of gene functions in hub modules were performed using the "clusterProfiler" R package (Wu et al., 2021) with the default parameters.

Identification of DEGs in HC, inactive and active SLE

The "DESeq2" R package was used to identify DEGs among the three groups with the threshold of $|\log_2 FC| > 1$ and $p < 0.05$. Besides, the "TCseq" R package "timeclust" function was used to characterize the gene expression patterns in three groups following the manuscript (DOI: 10.18129/B9.bioc.TCseq). The clusters that showed positive or negative relations with SLE activity were selected.

Identification of key genes

The Jvenn web tool was used to identify the overlapped intersection of the genes identified by WGCNA, time course sequencing (TCseq) and DEG analysis. These key genes were selected to perform principal component analysis (PCA) and heatmap analysis for validation.

Identification of potential biomarkers of SLE

First, we analyzed the key genes identified above with the "clusterProfiler" R package (Wu et al., 2021), and identified the genes that participated in regulation of top GO terms related to the development of SLE. Then, we compared the gene expression between HC and SLE patients with different disease activity and identified the potential biomarkers of diagnostic and therapeutic value for SLE.

Statistical analysis

All the statistical process and analysis in this manuscript were performed with R software, and $p < 0.05$ was considered as statistically significant.

Results

Gene expression profiles in HC, inactive and active SLE

Herein, 44 isolated $CD4^+$ T cells samples obtained from the dataset GSE97263 were processed. A total of 41,092 Ensemble IDs were converted into gene symbols. As the genes with average raw reads less than 1 were removed, 16,623 genes were selected for the following analysis. After data processing, we examined the gene expression differences in $CD4^+$ T cell isolated from HC and SLE patients with different disease activity (Figure 1). According to the PCA results, the gene signatures of these three groups could not be clearly separated (Figure 2A). Next, we analyzed the differences between HC and active SLE group, and found 834 upregulated genes and 252 downregulated genes, with the cut-off criteria of $p < 0.05$ and $|\log_2 FC| > 1$ (Figure 2B). Similarly, the volcano plots showed the difference of DEGs between HC and inactive SLE (Figure 2C), and inactive and active SLE group (Figure 2D). Next, the genes with the largest variance of gene expression were selected for heatmapping, as shown in the Supplementary Figure S1, indicating dynamic changes in gene expression during the development of SLE.

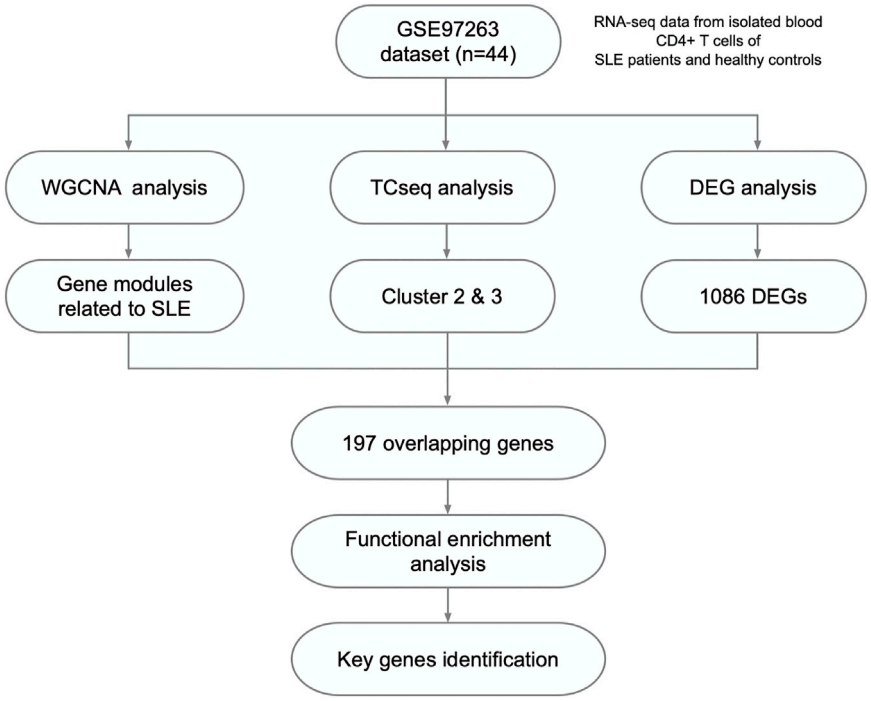


FIGURE 1
Study design and the workflow of this study.

Identification of hub modules through WGCNA

With no outlier samples, we calculated the optimal soft threshold power (β value) was 8, which was verified by scale-free topology analysis with $R^2 = 0.85$ (Supplementary Figures S2A,B). After merging similar modules with the cut-off value = 0.25, 14 modules from the weighted co-expression network were identified based on all the 16,623 genes (Supplementary Figures S2C,D). The gene numbers and detailed symbols in each module were shown in Supplementary Figure S2E and Supplementary Table S1, respectively. In order to explore the relationships among the above-mentioned modules, we quantified the module similarity by eigengene correlation, and the TOM heatmap showed strong correlation within the module groups (Supplementary Figures S2F).

Correlation between modules of interest and clinical traits

Next, identifying modules most associated with the disease activity is of great biological significance for biomarker development. According to the module-trait relationships in Figure 3A, the MElightcyan and MEsalmon modules were negatively related to disease activity (Cor = −0.45, $p = 0.002$ for MElightcyan, and Cor = −0.43, $p = 0.004$ for MEsalmon), while the MEcyan and MEbrown modules displayed positive relationship with disease activity (Cor = 0.7, $p = 1*10^{-7}$ for MEcyan, and Cor = 0.69, $p = 3*10^{-4}$ for MEbrown). Thus, these modules were selected for downstream analysis. As shown in Figures 3B–E, GS and MM were highly correlated, illustrating that genes significantly associated with disease activity were also the central elements of modules highly associated with this trait.

Functional analysis of hub modules

To investigate the correlated biological processes, GO enrichment analysis was carried out on all matched genes in these modules (Supplementary Table S2). In the salmon module, the enriched biological processes were mainly associated with nuclear-transcribed mRNA catabolic process (nonsense mediated decay), cotranslational protein targeting to membrane, SRP-dependent cotranslational protein targeting to membrane and translational initiation. In the cyan module, the biological processes were mainly enriched in the regulation of type I IFN production and response to IFN-α (Figure 3F).

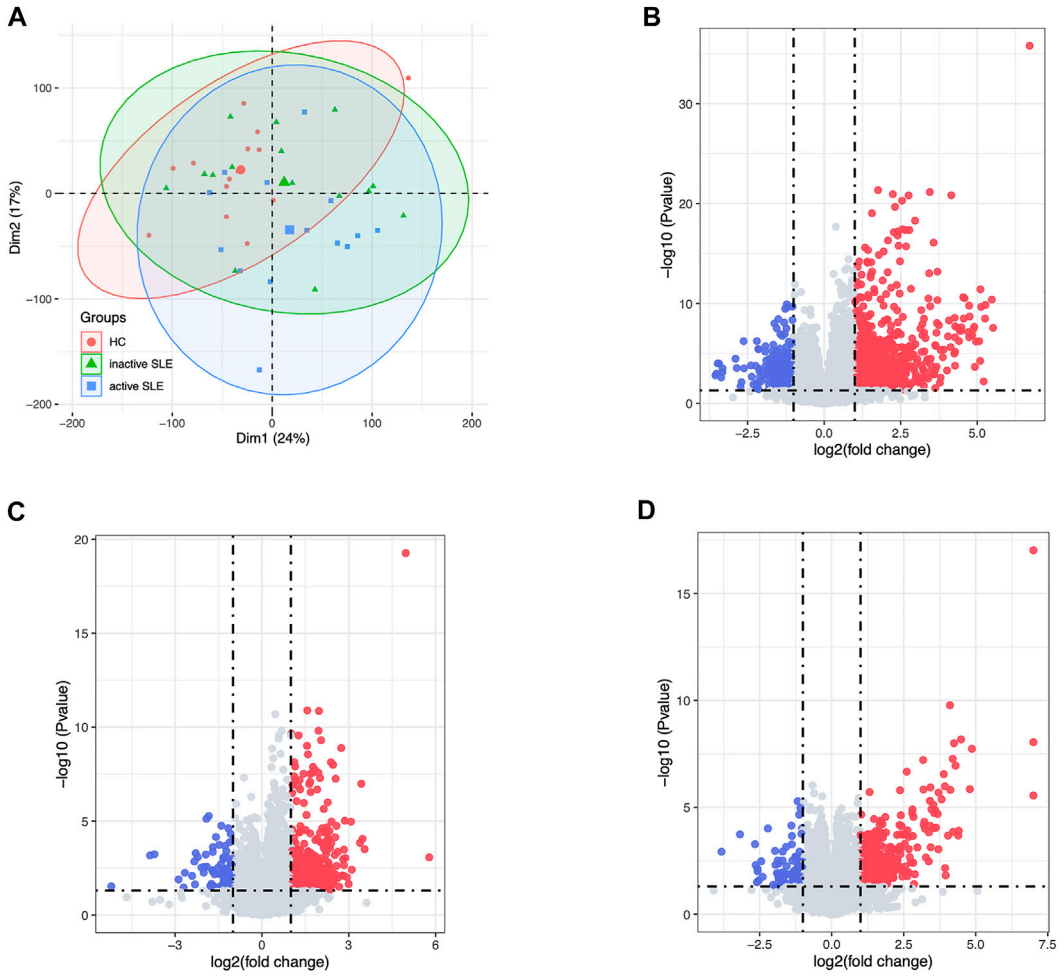


FIGURE 2
The DEG analysis showed the differences of gene expression profiles in HC, inactive and active SLE patients. **(A)** The PCA result of the gene expression of CD4$^+$ T cells in HC, inactive and active SLE patients. **(B–D)** The volcano plot of DEGs between HC and active SLE **(B)**, HC and inactive SLE **(C)**, and inactive SLE and active SLE **(D)**.

Identification of gene sets related to SLE disease activity

The above WGCNA results showed that the gene expression patterns of hub modules were significantly correlated with disease activity. To further clarify the gene sets closely related to SLE, TCseq analysis was used to analyze CD4$^+$ T cells from SLE patients with different disease activity. The detailed gene symbols in different clusters were listed in Supplementary Table S3. The results showed that the Cluster2 was positively correlated with SLE disease activity, while the Cluster3 was negatively related to the disease activity (Figure 4A). In KEGG enrichment analysis, we found that genes in the Cluster2, which were upregulated in SLE patients, were enriched in the pathways of cell cycle, necroptosis and p53 signaling. The GO analysis showed that these genes were enriched in T cell activation, DNA replication, regulation of innate immune response, type I IFN production, response to type I IFN and T cell migration (Figure 4B). In the Cluster3, the KEGG analysis showed that the downregulated genes were enriched in multiple signaling pathways, including Wnt and transforming growth factor-beta (TGF-β) signaling pathways. The GO analysis also revealed that these genes were related to positive regulation of cell projection organization, canonical Wnt signaling pathway, positive regulation of growth and positive regulation of protein binding (Figure 4C).

Based on combined analysis of WGCNA, TCseq and DEG, 146 genes showed positive correlation with disease activity and 51 genes showed negative correlation with disease activity in patients (Figures 5A,B). Therefore, these 197 genes could be used to distinguish active and inactive SLE patients (Figures

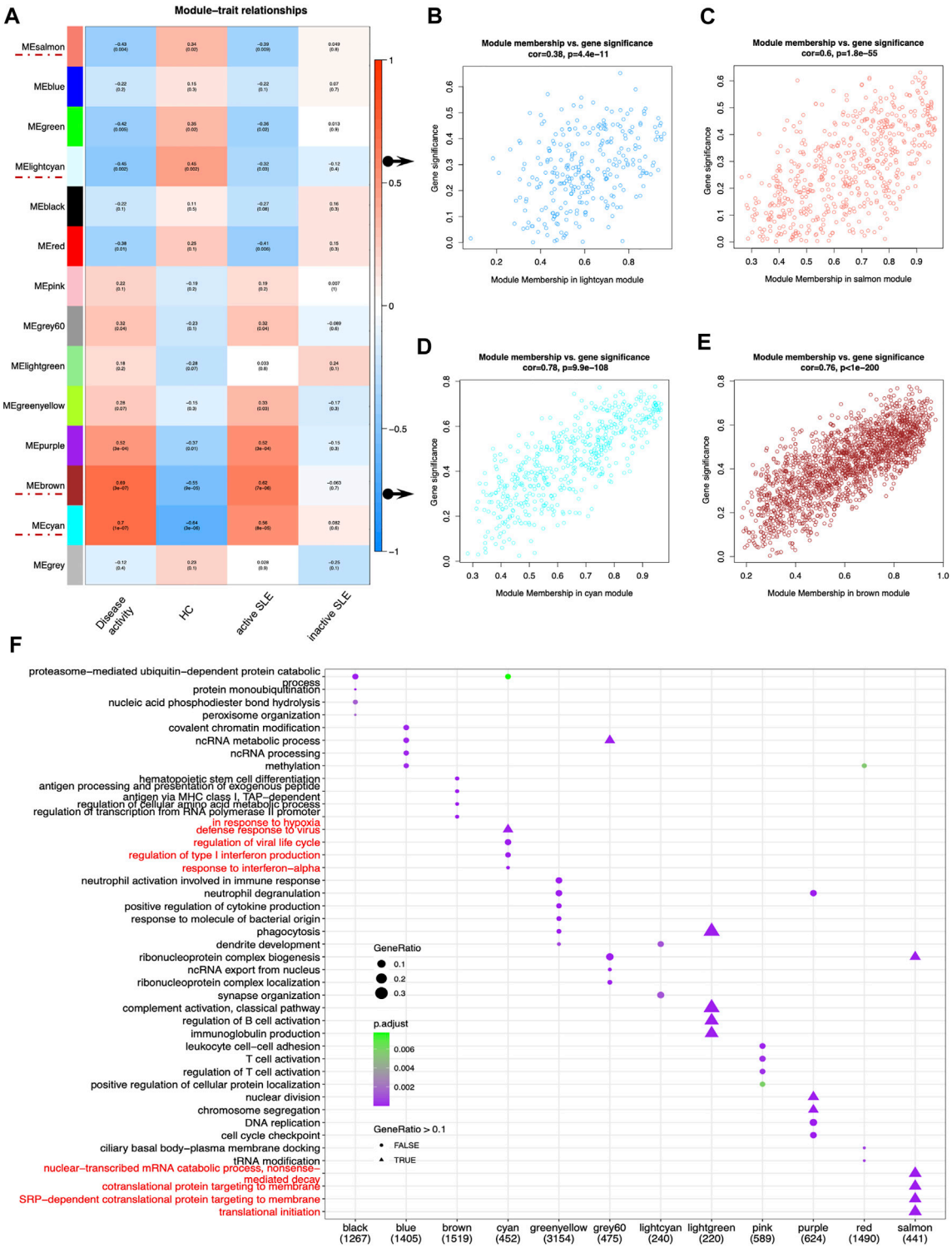


FIGURE 3
Main findings in the module-trait correlations through WGCNA. **(A)** Module-trait associations. Each row corresponded to a module eigengene (ME), while each column corresponded to a trait. Each cell contained the corresponding correlation and *p* value. The cells were color-coded by correlation according to the color legend. **(B–E)** The scatterplots of Gene Significance (GS) for disease activity *vs*. Module Membership (MM) in the lightcyan **(B)**, salmon **(C)**, cyan **(D)** and brown module **(E)**, which represented significant correlations between GS and MM in these modules. **(F)** The dot plot of GO enrichment analysis of the genes in different modules.

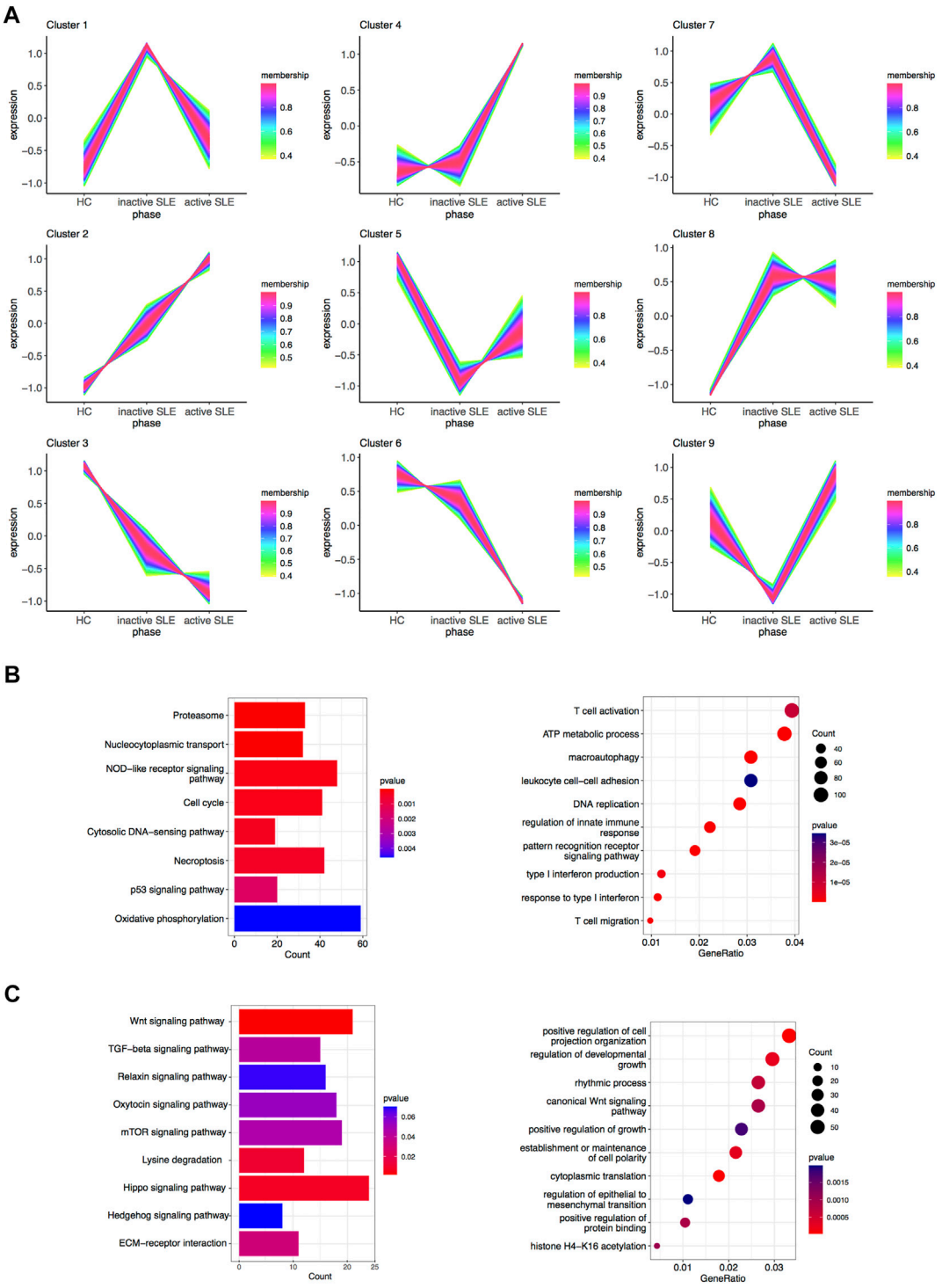


FIGURE 4
Identification of gene sets related to SLE disease activity. **(A)** Clustering of the gene expression patterns in CD4$^+$ T cells from HC, inactive and active SLE patients by TCseq analysis. A total of 9 clusters were obtained and the Cluster 2 and 3 were found significantly correlated with SLE disease activity. **(B)** KEGG (left) and GO (right) analysis of genes in the Cluster 2. **(C)** KEGG (left) and GO (right) analysis of genes in the Cluster 3.

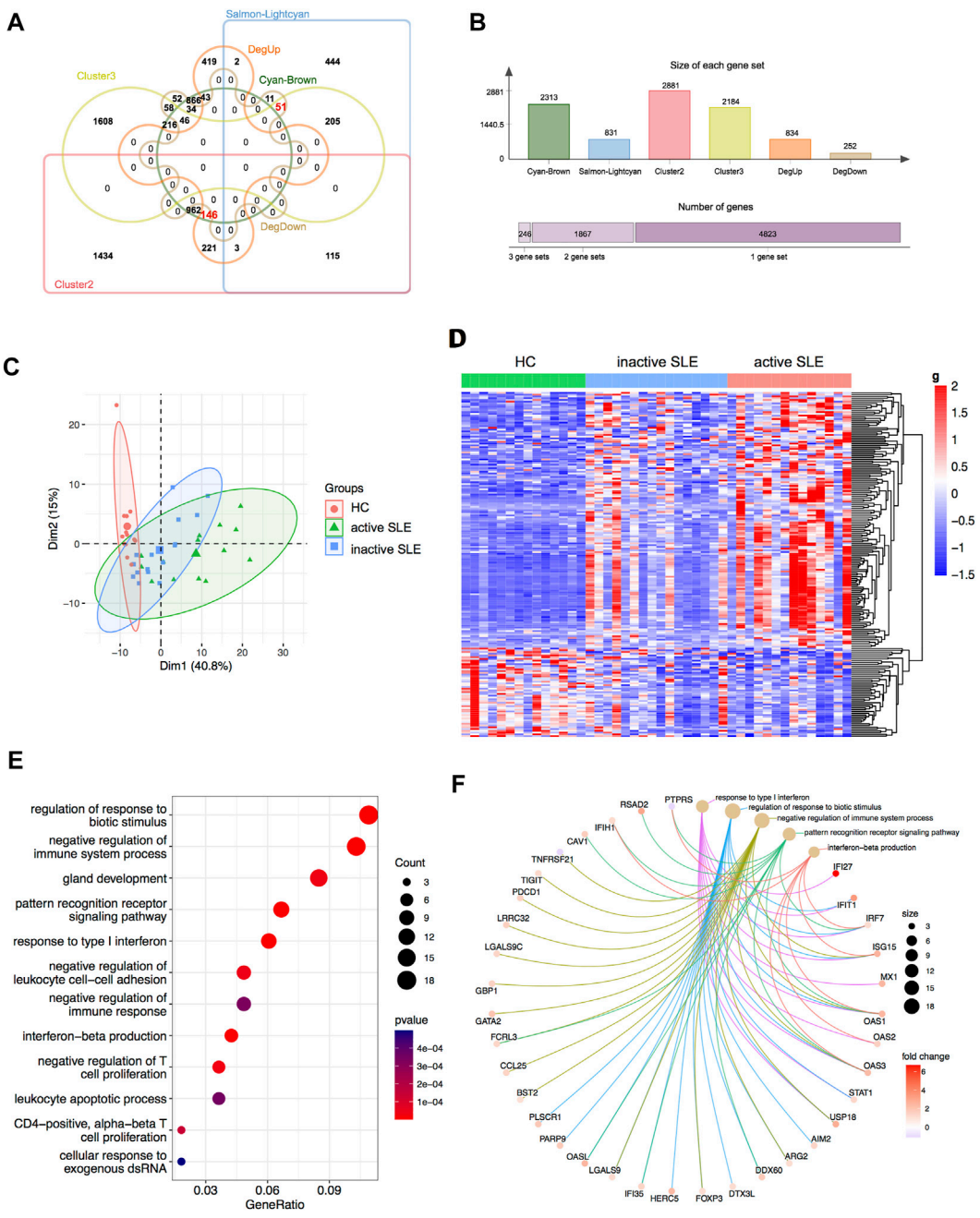


FIGURE 5
Identification of the candidate key genes. **(A,B)** The venn diagram **(A)** and bar plot **(B)** showed the number of genes in the salmon-lightcyan and cyan-brown modules, Cluster 2 and 3, upregulated and downregulated DEGs gene sets. **(C,D)** The PCA **(C)** and heatmap **(D)** analysis of the selected genes in HC, inactive and active SLE patients. **(E,F)** Dot plot **(E)** and network plot **(F)** of enriched terms of the candidate key genes through GO analysis.

5C,D). We furtherly analyzed the biological functions of these identified genes and found that they were enriched in the response to type I IFN, IFN-β production, negative regulation of T cell proliferation and immune system process (Figures 5E,F).

Identification of the potential biomarkers for SLE

The genes upregulated in SLE patients mainly focused on type I IFN response. We found 24 ISGs (*RSAD2*, *IFIT3*, *APOBEC3A*, *IFIT2*,

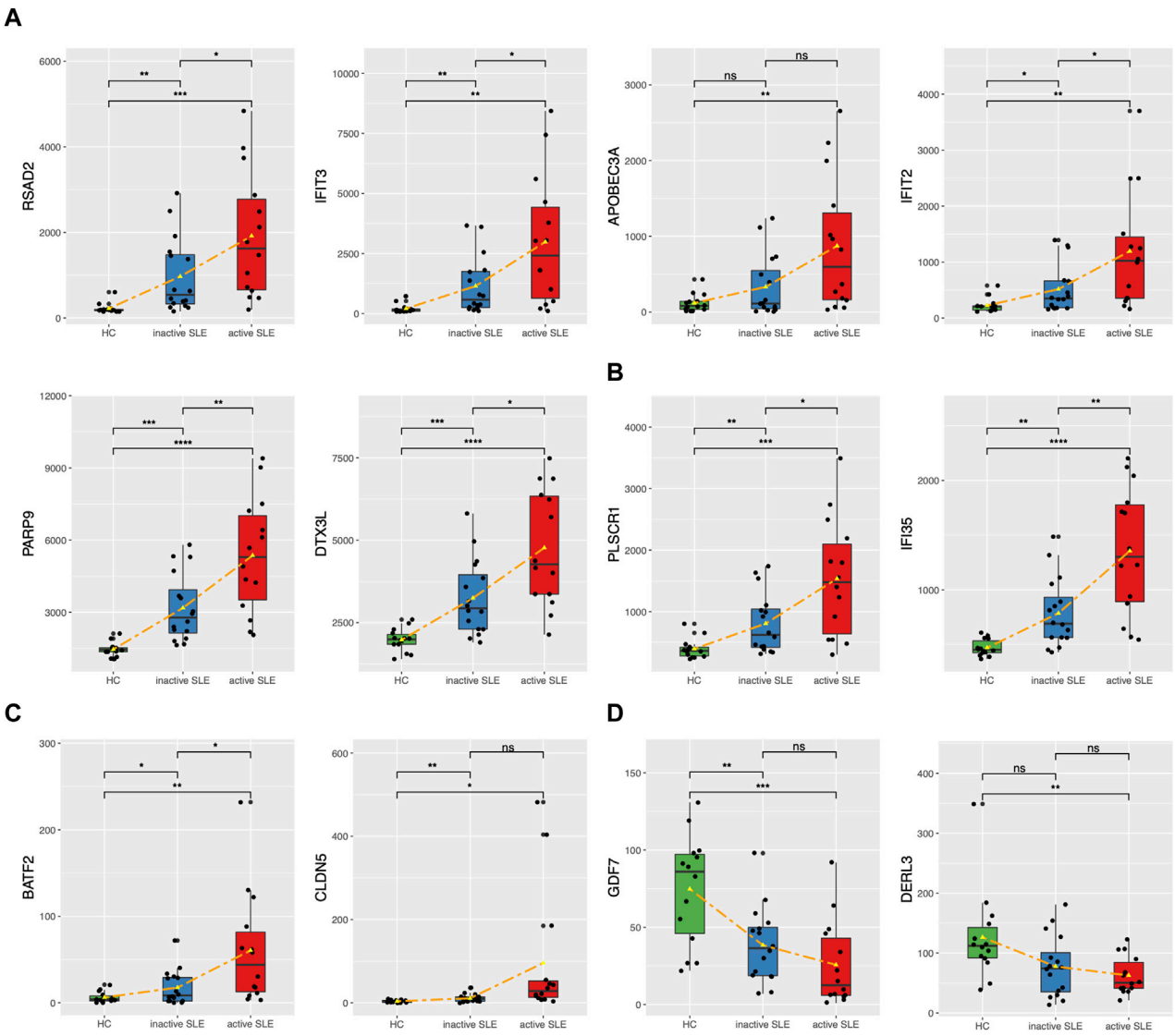


FIGURE 6
The expression levels of potential key genes in HC and SLE patients. **(A,B)** The expression levels of candidate ISGs in $CD4^+$ T cells from HC, inactive and active SLE patients. **(C,D)** The expression levels of upregulated BATF2 and CLDN5 **(C)** and downregulated GDF7 and DERL3 **(D)** in $CD4^+$ T cells from HC, inactive and active SLE patients. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$.

PARP9, *DTX3L*, *PLSCR1*, *IFI35*, *ISG15*, *CMPK2*, *HERC5*, *GBP1*, *IFI27*, *STAT1*, *MX1*, *IRF7*, *OAS1*, *OAS2*, *OAS3*, *OASL*, *IFIH1*, *SIGLEC1*, *LGALS9* and *BST2*) which were significantly associated with SLE disease activity (Figures 6A,B, Supplementary Figure S3). Among them, *PLSCR1* and *IFI35* were identified as the newfound crucial ISGs. Besides, *BATF2* and *CLDN5* were also identified as key genes (Figure 6C). The expression levels of the above genes were significantly upregulated, especially in those active SLE patients. On the other hand, the expression of *GDF7* and *DERL3* showed significant decrease in active or inactive SLE, compared with that in HC group (Figure 6D). The relevant research in other diseases also pointed that the 4 upregulated genes (*PLSCR1*, *IFI35*, *BATF2* and *CLDN5*) and 2 downregulated genes (*GDF7* and *DERL3*) exhibited immunoregulatory functions. Therefore, these 6 newfound genes may serve as potential biomarkers of SLE.

Discussion

It has been reported that lupus $CD4^+$ T cells had altered signaling and function, and the hyperactivation of these cells was

an important molecular feature of SLE patients. In this study, we analyzed the expression patterns of $CD4^+$ T cell in HC, inactive and active SLE through multiple analysis methods, and explored the molecular indicators for potential diagnostic biomarkers and therapeutic targets of SLE.

From methodology aspect, DEG analysis focuses on the differentially expressed genes among different groups, while WGCNA focuses on the correlations between the co-expression modules and the phenotypic and clinical traits, not merely the differences in gene expression profiles. TCseq can be applied for differential analysis between different time points and temporal pattern analysis and visualization of sequencing data. These three bioinformatical tools are complementary to one another to describe key relevant patterns to expanding our capacity for identifying novel biomarkers. In this study, through WGCNA analysis, we established gene expression module-disease activity relationship and found the main functional enrichment in the cyan module included regulation of type I IFN production and the IFN response. This observation is consistent with the current knowledge that ISGs are highly related with SLE disease activity (Crow, 2014; Gkirtzimanaki et al., 2018; Buang et al., 2021). From the perspective of treatment, monoclonal antibodies such as Anifrolumab (Anderson and Furie, 2020; Tanaka and Tummala, 2021) and Sifalimumab (Greth et al., 2017), which block the activation of type I IFNs, have demonstrated significant effectiveness in achieving the composite endpoints in active SLE patients. The medical researchers have also tried to treat lupus by inhibiting type I IFNs in a variety of ways, such as glucocorticoids (Kirou and Gkrouzman, 2013), nicotinamide riboside (Wu et al., 2022) and mTOR inhibitor (Murayama et al., 2020). On the other hand, the salmon module was enriched in biological processes of translational initiation and cotranslational protein targeting to membrane, showing that the expression level negatively correlated with disease activity. The major consequence of perturbing cotranslational targeting for disease progression is so far largely unexploited. We speculate that it may be involved in T cell differentiation, which accompanied by the expression and secretion of a large number of cytokines, as well as communication between cells.

Furtherly, by combining the TCseq with DEG analysis, we identified the presence of 197 genes as closely correlated with disease activity in patients, that may play an integral role in the development of SLE. GO analysis showed that DEGs were mainly involved in negative regulation of immune system process, the response to type I IFN, IFN-β production and negative regulation of T cell proliferation. A total of 24 ISGs showed significant upregulation in SLE patients, especially in those who presented with higher disease activity. The results are consistent with the phenomenon of over-activated IFN response in lupus patients, and further confirm that ISGs play a vital role in the pathogenesis of SLE. Among these ISGs, we identified *PLSCR1* and *IFI35* as the newfound crucial genes. *PLSCR1* is a member of the phospholipid scramblases protein family and involved in regulating phospholipid movements within the plasma membrane. Several reports found that significant hypomethylation of differentially methylated sites in SLE was associated with *PLSCR1*(Yeung et al., 2017; Joseph et al., 2019; He et al., 2022). Besides, elevated expression of *PLSCR1* was found in monocytes from SLE patients (Suzuki et al., 2010), and it was also involved in the modulation of the phagocytic process in differentiated macrophages (Herate et al., 2016). *IFI35* reflects the type I IFN activity induced through the JAK-STAT phosphorylation (De Masi et al., 2021). Elevated expression levels of *IFI35* were found in serum of LN patients, which promoted LPS-caused inflammatory response and cell apoptosis (Zhang et al., 2021). *IFI35* also showed regulatory effects on multiple immune cells by activating macrophages and dendritic cells and promoting naïve T cell differentiation into Th1 and Th17 cells (Xiahou et al., 2017; Jing et al., 2021). A significant elevation in *IFI35* expression in active SLE was also found in our study, indicating that *IFI35* may be associated with the dysregulation of host IFN production and immune cell function in SLE.

Besides, we also explored upregulated *BATF2* and *CLDN5* as critical genes in SLE. *BATF2* was significantly induced and involved in gene regulation of IFN-γ-activated classical macrophages (Roy et al., 2015), and inhibited Th17 responses by suppressing IL-23a expression (Kitada et al., 2017; Kayama et al., 2019). In SLE, *in vitro* experiments indicated *BATF2* may be involved in the impairment of translational and proliferative responses to mitogens in T cells (Ge et al., 2021). *CLDN5* encodes tight junction protein and plays a role in C5a/C5aR1 signaling, which was reported to be related to the impaired brain-blood-barrier (BBB) in SLE with neurological complications (Mahajan et al., 2015). These abnormal expression patterns may eventually disturb immune function with $CD4^+$ T cells activation and proliferation. Among these downregulated genes, we identified *GDF7* and *DERL3* as key genes. *GDF7* encodes a secreted ligand of the TGF-β superfamily of proteins. Recent research showed that *GDF7* can exhibit positive regulatory effects on Tregs *via* increasing the expression of *FOXP3* and *CTLA4* (Ding et al., 2021). Thus, the downregulation of *GDF7* in $CD4^+$ T cells may lead to impaired suppressive functions of lupus Tregs. *DERL3* encodes proteins belong to the derlin family, which resides in the endoplasmic reticulum (ER). Recent reports using WGCNA or RNA-seq suggested that the function of DERL3 may correlate with plasma cells (Gao et al., 2022; da Silva et al., 2020). Since $CD4^+$ T cells act as main helper cells for plasma cell production and cytokines secretion, we suggest that the decreased expression of *DERL3* in lupus $CD4^+$ T cells may contribute to the pathogenesis of SLE. Taken together, our data indicated the potential diagnostic and therapeutic value of *GDF7* and *DERL3* in SLE.

There are still several limitations in this study. First of all, the diagnostic or prognostic value of these key genes require a large number of blood samples for validation. Secondly, we did not further clarify the molecular mechanisms of the identified genes. Finally, although we have performed a detailed bioinformatics analysis, some vital genes in the pathogenesis and progression of SLE may still be missed. Thus, further analysis and detailed experiments are needed to definitely establish the predictive biomarkers and explicitly evaluate the performance.

In summary, based on integrated bioinformatical analysis, we found gene sets highly related to SLE disease activity. Besides, some crucial genes mediating the development of SLE were identified, including some previously reported ISGs. In particular, we found 4 upregulated genes (*PLSCR1*, *IFI35*, *BATF2* and *CLDN5*) and downregulated expression of *GDF7* and *DERL3* in SLE patients. Therefore, our findings identified 6 novel potential biomarkers in lupus $CD4^+$ T cells, which provided new insights into the development and treatment of SLE. The underlying mechanisms of these genes are still need to be further explored.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE97263.

Author contributions

ZL, ZW and LS contributed to conception and design of the study. ZL and ZW organized the database, performed the statistical analysis and wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This work was supported by National Key R&D Program of China (grant no. 2020YFA0710800), the Key Program of National Natural Science Foundation of China (grant no. 81930043) and the Major International (Regional) Joint Research Project of China (grant no. 81720108020).

Acknowledgments

The authors thank Xin Wen and Dr. Wei Shen for the comments on the data analysis, Dr. Dandan Wang and Dr. Linyu Geng for the English revision of this manuscript.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.941221/full#supplementary-material

References

Anderson, E., and Furie, R. (2020). Anifrolumab in systemic lupus erythematosus: Current knowledge and future considerations. *Immunotherapy* 12 (5), 275–286. doi:10.2217/imt-2020-0017

Buang, N., Tapeng, L., Gray, V., Sardini, A., Whilding, C., Lightstone, L., et al. (2021). Type I interferons affect the metabolic fitness of CD8(+) T cells from patients with systemic lupus erythematosus. *Nat. Commun.* 12 (1), 1980. doi:10.1038/s41467-021-22312-y

Chen, Z., Lan, R., Ye, K., Chen, H., Chen, C., Xu, Y., et al. (2021). Prioritization of diagnostic and prognostic biomarkers for lupus nephritis based on integrated bioinformatics analyses. *Front. Bioeng. Biotechnol.* 9, 717234. doi:10.3389/fbioe.2021.717234

Crow, M. K. (2014). Advances in understanding the role of type I interferons in systemic lupus erythematosus. *Curr. Opin. Rheumatol.* 26 (5), 467–474. doi:10.1097/BOR.0000000000000087

da Silva, F. A. R., Pascoal, L. B., Dotti, I., Setsuko Ayrizono, M. L., Aguilar, D., Rodrigues, B. L., et al. (2020). Whole transcriptional analysis identifies markers of B, T and plasma cell signaling pathways in the mesenteric adipose tissue associated with Crohn's disease. *J. Transl. Med.* 18 (1), 44. doi:10.1186/s12967-020-02220-3

De Masi, R., Orlando, S., Bagordo, F., and Grassi, T. (2021). IFP35 is a relevant factor in innate immunity, multiple sclerosis, and other chronic inflammatory diseases: A review. *Biol. (Basel)* 10 (12), 1325. doi:10.3390/biology10121325

Ding, M., Malhotra, R., Ottosson, T., Lundqvist, M., Mebrahtu, A., Brengdahl, J., et al. (2021). Secretome screening reveals immunomodulating functions of IFNα-7, PAP and GDF-7 on regulatory T-cells. *Sci. Rep.* 11 (1), 16767. doi:10.1038/s41598-021-96184-z

Dorner, T., and Furie, R. (2019). Novel paradigms in systemic lupus erythematosus. *Lancet* 393 (10188), 2344–2358. doi:10.1016/S0140-6736(19)30546-X

Feng, X., Huang, J., Liu, Y., Xiao, L., Wang, D., Hua, B., et al. (2015). Identification of interferon-inducible genes as diagnostic biomarker for systemic lupus erythematosus. *Clin. Rheumatol.* 34 (1), 71–79. doi:10.1007/s10067-014-2799-4

Gao, X., Jiang, C., Yao, S., Ma, L., Wang, X., Cao, Z., et al. (2022). Identification of hub genes related to immune cell infiltration in periodontitis using integrated bioinformatic analysis. *J. Periodontal Res.* 57 (2), 392–401. doi:10.1111/jre.12970

Ge, L., Zhang, Y., Zhao, X., Wang, J., Zhang, Y., Wang, Q., et al. (2021). EIF2AK2 selectively regulates the gene transcription in immune response and histones associated with systemic lupus erythematosus. *Mol. Immunol.* 132, 132–141. doi:10.1016/j.molimm.2021.01.030

Gkirtzimanaki, K., Kabrani, E., Nikoleri, D., Polyzos, A., Blanas, A., Sidiropoulos, P., et al. (2018). IFNα impairs autophagic degradation of mtDNA promoting autoreactivity of SLE monocytes in a STING-dependent fashion. *Cell. Rep.* 25 (4), 921–933. doi:10.1016/j.celrep.2018.09.001

Goulielmos, G. N., Zervou, M. I., Vazgiourakis, V. M., Ghodke-Puranik, Y., Garyfallos, A., Niewold, T. B., et al. (2018). The genetics and molecular pathogenesis of systemic lupus erythematosus (SLE) in populations of different ancestry. *Gene* 668, 59–72. doi:10.1016/j.gene.2018.05.041

Greth, W., Robbie, G. J., Brohawn, P., Hultquist, M., and Yao, B. (2017). Targeting the interferon pathway with sifalimumab for the treatment of systemic lupus erythematosus. *Immunotherapy* 9 (1), 57–70. doi:10.2217/imt-2016-0090

He, Z., Zhou, S., Yang, M., Zhao, Z., Mei, Y., Xin, Y., et al. (2022). Comprehensive analysis of epigenetic modifications and immune-cell infiltration in tissues from patients with systemic lupus erythematosus. *Epigenomics* 14 (2), 81–100. doi:10.2217/epi-2021-0318

Herate, C., Ramdani, G., Grant, N. J., Marion, S., Gasman, S., Niedergang, F., et al. (2016). Phospholipid scramblase 1 modulates FcR-mediated phagocytosis in differentiated macrophages. *PLoS One* 11 (1), e0145617. doi:10.1371/journal.pone.0145617

Jang, E., Cho, S., Pyo, S., Nam, J. W., and Youn, J. (2021). An inflammatory loop between spleen-derived myeloid cells and CD4(+) T cells leads to accumulation of long-lived plasma cells that exacerbates lupus autoimmunity. *Front. Immunol.* 12, 631472. doi:10.3389/fimmu.2021.631472

Jing, X., Yao, Y., Wu, D., Hong, H., Feng, X., Xu, N., et al. (2021). IFP35 family proteins promote neuroinflammation and multiple sclerosis. *Proc. Natl. Acad. Sci. U. S. A.* 118 (32), e2102642118. doi:10.1073/pnas.2102642118

Joseph, S., George, N. I., Green-Knox, B., Treadwell, E. L., Word, B., Yim, S., et al. (2019). Epigenome-wide association study of peripheral blood mononuclear cells in systemic lupus erythematosus: Identifying DNA methylation signatures associated with interferon-related genes based on ethnicity and SLEDAI. *J. Autoimmun.* 96, 147–157. doi:10.1016/j.jaut.2018.09.007

Kayama, H., Tani, H., Kitada, S., Opasawatchai, A., Okumura, R., Motooka, D., et al. (2019). BATF2 prevents T-cell-mediated intestinal inflammation through regulation of the IL-23/IL-17 pathway. *Int. Immunol.* 31 (6), 371–383. doi:10.1093/intimm/dxz014

Kirou, K. A., and Gkrouzman, E. (2013). Anti-interferon alpha treatment in SLE. *Clin. Immunol.* 148 (3), 303–312. doi:10.1016/j.clim.2013.02.013

Kitada, S., Kayama, H., Okuzaki, D., Koga, R., Kobayashi, M., Arima, Y., et al. (2017). BATF2 inhibits immunopathological Th17 responses by suppressing Il23a expression during Trypanosoma cruzi infection. *J. Exp. Med.* 214 (5), 1313–1331. doi:10.1084/jem.20161076

Langfelder, P., and Horvath, S. (2008). Wgcna: an R package for weighted correlation network analysis. *BMC Bioinforma.* 9, 559. doi:10.1186/1471-2105-9-559

Liu, X., Ding, Y., Zheng, X., Huang, H., Shi, L., Yang, X., et al. (2021). Small RNAs encoded by human endogenous retrovirus K overexpressed in PBMCs may contribute to the diagnosis and evaluation of systemic lupus erythematosus as novel biomarkers. *Hum. Mol. Genet.* 31, 1407–1416. doi:10.1093/hmg/ddab32

Luo, S., Long, H., and Lu, Q. (2020). Recent advances in understanding pathogenesis and therapeutic strategies of Systemic Lupus Erythematosus. *Int. Immunopharmacol.* 89, 107028. doi:10.1016/j.intimp.2020.107028

Mahajan, S. D., Parikh, N. U., Woodruff, T. M., Jarvis, J. N., Lopez, M., Hennon, T., et al. (2015). C5a alters blood-brain barrier integrity in a human *in vitro* model of systemic lupus erythematosus. *Immunology* 146 (1), 130–143. doi:10.1111/imm.12489

Mak, A., and Kow, N. Y. (2014). The pathology of T cells in systemic lupus erythematosus. *J. Immunol. Res.* 2014, 419029. doi:10.1155/2014/419029

Moulton, V. R., and Tsokos, G. C. (2011). Abnormalities of T cell signaling in systemic lupus erythematosus. *Arthritis Res. Ther.* 13 (2), 207. doi:10.1186/ar3251

Murayama, G., Chiba, A., Kuga, T., Makiyama, A., Yamaji, K., Tamura, N., et al. (2020). Inhibition of mTOR suppresses IFNα production and the STING pathway in monocytes from systemic lupus erythematosus patients. *Rheumatol. Oxf.* 59 (10), 2992–3002. doi:10.1093/rheumatology/keaa060

Obermoser, G., and Pascual, V. (2010). The interferon-alpha signature of systemic lupus erythematosus. *Lupus* 19 (9), 1012–1019. doi:10.1177/0961203310371161

Postal, M., Vivaldo, J. F., Fernandez-Ruiz, R., Paredes, J. L., Appenzeller, S., Niewold, T. B., et al. (2020). Type I interferon in the pathogenesis of systemic lupus erythematosus. *Curr. Opin. Immunol.* 67, 87–94. doi:10.1016/j.coi.2020.10.014

Ronnblom, L. (2016). The importance of the type I interferon system in autoimmunity. *Clin. Exp. Rheumatol.* 34 (98), 21–24.

Roy, S., Guler, R., Parihar, S. P., Schmeier, S., Kaczkowski, B., Nishimura, H., et al. (2015). Batf2/Irf1 induces inflammatory responses in classically activated macrophages, lipopolysaccharides, and mycobacterial infection. *J. Immunol.* 194 (12), 6035–6044. doi:10.4049/jimmunol.1402521

Shen, L., Lan, L., Zhu, T., Chen, H., Gu, H., Wang, C., et al. (2021). Identification and validation of IFI44 as key biomarker in lupus nephritis. *Front. Med.* 8, 762848. doi:10.3389/fmed.2021.762848

Sun, G., Zhu, P., Dai, Y., and Chen, W. (2019). Bioinformatics analysis of the core genes related to lupus nephritis through a network and pathway-based approach. *DNA Cell. Biol.* 38 (7), 639–650. doi:10.1089/dna.2019.4631

Suzuki, E., Amengual, O., Atsumi, T., Oku, K., Hashimoto, T., Kataoka, H., et al. (2010). Increased expression of phospholipid scramblase 1 in monocytes from patients with systemic lupus erythematosus. *J. Rheumatol.* 37 (8), 1639–1645. doi:10.3899/jrheum.091420

Tanaka, Y., and Tummala, R. (2021). Anifrolumab, a monoclonal antibody to the type I interferon receptor subunit 1, for the treatment of systemic lupus erythematosus: An overview from clinical trials. *Mod. Rheumatol.* 31 (1), 1–12. doi:10.1080/14397595.2020.1812201

Teruel, M., and Alarcon-Riquelme, M. E. (2016). The genetic basis of systemic lupus erythematosus: What are the risk factors and what have we learned. *J. Autoimmun.* 74, 161–175. doi:10.1016/j.jaut.2016.08.001

Tsokos, G. C. (2011). Systemic lupus erythematosus. *N. Engl. J. Med.* 365 (22), 2110–2121. doi:10.1056/NEJMra1100359

Wu, J., Singh, K., Lin, A., Meadows, A. M., Wu, K., Shing, V., et al. (2022). Boosting NAD+ blunts toll-like receptor-4 induced type-I interferon in control and systemic lupus erythematosus monocytes. *J. Clin. Invest.* 132, e139828. doi:10.1172/JCI139828

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innov. (N Y).* 2 (3), 100141. doi:10.1016/j.xinn.2021.100141

Xiahou, Z., Wang, X., Shen, J., Zhu, X., Xu, F., Hu, R., et al. (2017). NMI and IFP35 serve as proinflammatory DAMPs during cellular infection and injury. *Nat. Commun.* 8 (1), 950. doi:10.1038/s41467-017-00930-9

Yan, S., Wang, W., Gao, G., Cheng, M., Wang, X., Wang, Z., et al. (2018). Key genes and functional coexpression modules involved in the pathogenesis of systemic lupus erythematosus. *J. Cell. Physiol.* 233 (11), 8815–8825. doi:10.1002/jcp.26795

Yao, M., Gao, C., Zhang, C., Di, X., Liang, W., Sun, W., et al. (2020). Identification of molecular markers associated with the pathophysiology and treatment of lupus nephritis based on integrated transcriptome analysis. *Front. Genet.* 11, 583629. doi:10.3389/fgene.2020.583629

Yeung, K. S., Chung, B. H., Choufani, S., Mok, M. Y., Wong, W. L., Mak, C. C., et al. (2017). Genome-Wide DNA methylation analysis of Chinese patients with systemic lupus erythematosus identified hypomethylation in genes related to the type I interferon pathway. *PLoS One* 12 (1), e0169553. doi:10.1371/journal.pone.0169553

Yuan, S., Zeng, Y., Li, J., Wang, C., Li, W., He, Z., et al. (2022). Phenotypical changes and clinical significance of CD4(+)/CD8(+) T cells in SLE. *Lupus Sci. Med.* 9 (1), e000660. doi:10.1136/lupus-2022-000660

Zhang, L. H., Jiang, S. Z., Guo, X., Xiao, B., Li, Q., Chen, J. Y., et al. (2021). MiR-146b-5p targets IFI35 to inhibit inflammatory response and apoptosis *via* JAK1/STAT1 signalling in lipopolysaccharide-induced glomerular cells. *Autoimmunity* 54 (7), 430–438. doi:10.1080/08916934.2020.1864730

Zhao, M., Li, M. Y., Gao, X. F., Jia, S. J., Gao, K. Q., Zhou, Y., et al. (2018). Downregulation of BDH2 modulates iron homeostasis and promotes DNA demethylation in CD4(+) T cells of systemic lupus erythematosus. *Clin. Immunol.* 187, 113–121. doi:10.1016/j.clim.2017.11.002

Zhao, X., Zhang, L., Wang, J., Zhang, M., Song, Z., Ni, B., et al. (2021). Identification of key biomarkers and immune infiltration in systemic lupus erythematosus by integrated bioinformatics analysis. *J. Transl. Med.* 19 (1), 35. doi:10.1186/s12967-020-02698-x

frontiers | Frontiers in Genetics

MINI REVIEW
published: 01 September 2022
doi: 10.3389/fgene.2022.929162

Roles of AIM2 Gene and AIM2 Inflammasome in the Pathogenesis and Treatment of Psoriasis

Jieyi Wang[1,2,3†], *Jing Gao*[4,5†], *Cong Huang*[1,2], *Sohyun Jeong*[6,7], *Randy Ko*[8], *Xue Shen*[9], *Chaofeng Chen*[1,2], *Weilong Zhong*[1,2], *Yanfen Zou*[1,2], *Bo Yu*[1,2,3] *and Changbing Shen*[1,2*]

[1]Department of Dermatology, Peking University Shenzhen Hospital, Shenzhen, Guangdong, China, [2]Shenzhen Key Laboratory for Translational Medicine of Dermatology, Shenzhen Peking University—The Hong Kong University of Science and Technology Medical Center, Shenzhen, Guangdong, China, [3]School of Clinical Medicine, Health Science Center, Shenzhen University, Shenzhen, Guangdong, China, [4]Department of Dermatology, The Second Affiliated Hospital, Anhui Medical University, Hefei, Anhui, China, [5]Anhui Provincial Institute of Translational Medicine, Hefei, Anhui, China, [6]Marcus Institute for Aging Research at Hebrew SeniorLife, Boston, MA, United States, [7]Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, United States, [8]Department of Internal Medicine, University of New Mexico Health Sciences Center, Albuquerque, NM, United States, [9]Department of Dermatology, Chengdu Second People's Hospital, Chengdu, Sichuan, China

OPEN ACCESS

Edited by:
Yonghu Sun,
Shandong Provincial Hospital of Dermatology, China

Reviewed by:
Rui-qun Qi,
The First Hospital of China Medical University, China
Xiaojing Yu,
Qilu Hospital, Shandong University, China

***Correspondence:**
Changbing Shen
cambridge2008@126.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: 26 April 2022
Accepted: 22 June 2022
Published: 01 September 2022

Citation:
Wang J, Gao J, Huang C, Jeong S, Ko R, Shen X, Chen C, Zhong W, Zou Y, Yu B and Shen C (2022) Roles of AIM2 Gene and AIM2 Inflammasome in the Pathogenesis and Treatment of Psoriasis. Front. Genet. 13:929162. doi: 10.3389/fgene.2022.929162

Psoriasis is an immune-mediated chronic inflammatory skin disease caused by a combination of environmental incentives, polygenic genetic control, and immune regulation. The inflammation-related gene absent in melanoma 2 (*AIM2*) was identified as a susceptibility gene for psoriasis. AIM2 inflammasome formed from the combination of AIM2, PYD-linked apoptosis-associated speck-like protein (ASC) and Caspase-1 promotes the maturation and release of inflammatory cytokines such as IL-1β and IL-18, and triggers an inflammatory response. Studies showed the genetic and epigenetic associations between *AIM2* gene and psoriasis. *AIM2* gene has an essential role in the occurrence and development of psoriasis, and the inhibitors of AIM2 inflammasome will be new therapeutic targets for psoriasis. In this review, we summarized the roles of the *AIM2* gene and AIM2 inflammasome in pathogenesis and treatment of psoriasis, hopefully providing a better understanding and new insight into the roles of *AIM2* gene and AIM2 inflammasome in psoriasis.

Keywords: psoriasis, AIM2, AIM2 inflammasome, pathogenesis, treatment

1 INTRODUCTION

Psoriasis is an inflammatory skin disease caused by a combination of environmental incentives, polygenic genetic control, and immune regulation (Branisteanu et al., 2022). The worldwide average prevalence of psoriasis is 2% (Samotij et al., 2020), and the disease is often characterized by recurrence and incurability. It is often accompanied by complex comorbidities such as cardiovascular and autoimmune diseases (Gao et al., 2021), bringing great physical and mental harm to patients, and causing a substantial social–economic burden. The pathogenesis and novel therapies for psoriasis are hot research topics. In the past two decades, many susceptibility genes/loci and epigenetic modification factors of psoriasis have been identified (Nedoszytko et al., 2020). The inflammation-related gene absent in melanoma 2 (*AIM2*) was identified as a susceptibility gene for psoriasis (Zuo et al., 2015), and the function of *AIM2* gene and its role in psoriasis were explored

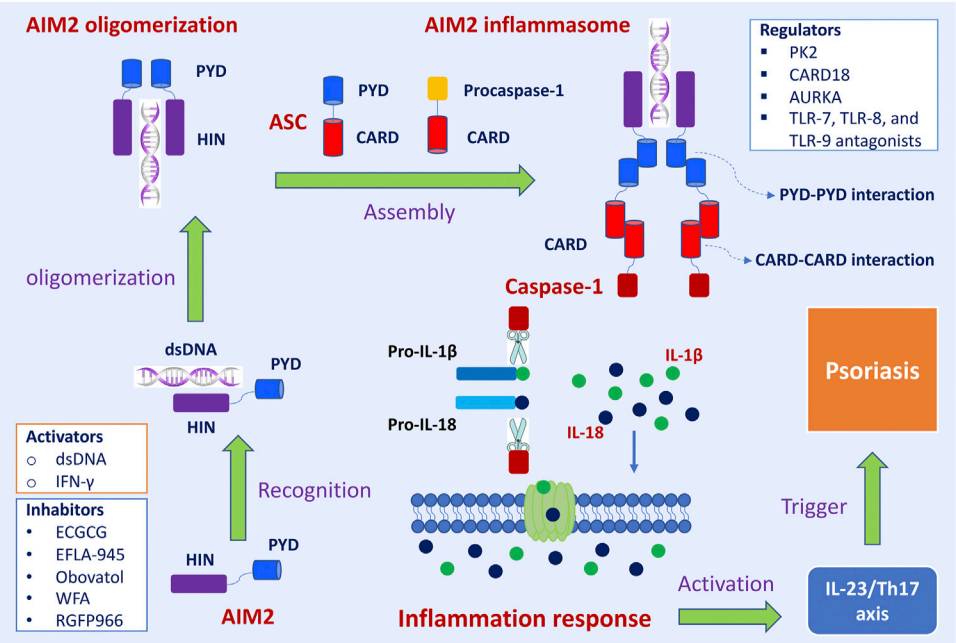


FIGURE 1 | The assembly and activation of AIM2 inflammasome and subsequently regulatory and trigger pathways in psoriasis. Upon sensing viral DNA, self-derived dsDNA, and cytosolic bacterium, the HIN domain of AIM2 directly recognizes dsDNA in a sequence-independent manner, which triggers the assembly of the AIM2 oligomerization. The PYD domain of AIM2 interacts with the PYD of a recruiting adapter protein ASC, resulting in a high polymer complex AIM2 inflammasome. Inactive procaspases-1 are recruited into AIM2 inflammasome via the CARD-CARD interaction. When the main components of the inflammasome are connecting and the active inflammasome is formed, it directly recruits and cleaves pro-caspase1 into active caspase-1, which proteolytically activates the pro-inflammatory cytokines IL-1β and IL-18. These inflammatory cytokines directly induce inflammatory responses and participate in the occurrence and development of diseases. The active IL-1β and IL-18 involve in IL23/Th17 pathway and then induce many kinds of chemokines and inflammatory cytokines, which trigger the development of psoriasis. Interferon-gamma (IFN-γ) promoted the expression of the *AIM2* gene. Epigallocatechin gallate (EGCG), EFLA-945, obovatol, withaferin A (WFA), and RGFP966 are inhibitory effects on AIM2. Prokineticin 2 (PK2), caspase recruitment domain family member 18 (CARD18), aurora kinase A (AURKA), TLR-7, TLR-8, and TLR-9 antagonists are regulators of AIM2 inflammasome signaling pathway.

(Ciążyńska et al., 2021; Liang et al., 2021). In this review, we systemically summarized genetic and epigenetic associations between *AIM2* gene and psoriasis, discussed the roles of *AIM2* gene and AIM2 inflammasome in the pathogenesis of psoriasis, and provided a better understanding of AIM2 inflammasome as a promising therapeutic target for psoriasis.

2 *AIM2* GENE AND AIM2 INFLAMMASOME

2.1 *AIM2* Gene

DeYoung et al. (1997) reported a novel gene, *AIM2*, also known as *PYHIN4*, located on chromosome 1q23.1-q23.2. The protein AIM2, encoded by the *AIM2* gene is a member of the IFI20X/IFI16 family, consisting of a C-terminal HIN domain and an N-terminal pyrin domain (PYD). RNA sequencing (RNA-seq) of 27 different human tissues showed that the expression level of *AIM2* gene in descending order are: lymph nodes, appendix, and spleen. *AIM2* gene is also expressed at a relatively low level in the skin compared with the top three tissues (Fagerberg et al., 2014). The expression of *AIM2* gene can be promoted by interferon-gamma (IFN-γ) (Tang H. et al., 2021). *AIM2* gene with various functions in the occurrence and development of diseases, the most common function is to initiate the assembly process of the AIM2 inflammasome. AIM2 response to dsDNA and then induce AIM2-dependent release of IL-18 and IL-1β, which plays a critical role as a trigger of autoimmune diseases, including psoriasis (Dombrowski et al., 2011), systemic lupus erythematosus (Shin et al., 2019), primary Sjogren's syndrome (Vakrakou et al., 2020). *AIM2* gene plays two-sided roles in tumorigenesis or anti-tumorigenesis in different tumors (Choubey et al., 2000). *AIM2* gene with the tumor-promoting effects in nonsmall-cell lung cancer (NSCLC) via the inflammasome-dependent manner and regulation of mitochondrial dynamics (Qi et al., 2020). On the contrary, *AIM2* gene is required to restrain the progression of colon cancer through an inflammasome-independent manner, proliferation control of intestinal stem cells, and the regulation of gut microbiota, suggesting that *AIM2* gene plays a protective role in colorectal cancer (Wilson et al., 2015; Zhang Z. et al., 2017).

2.2 AIM2 Inflammasome

In 2009, the composition of the AIM2 inflammasome was reported (Fernandes-Alnemri et al., 2009; Hornung et al., 2009). AIM2 can recognize double-stranded DNA (dsDNA) and interacts with N-terminal PYD. PYD-linked apoptosis-associated speck-like protein (ASC) induces the recruitment of Caspase-1 by the caspase recruitment domain (CARD) of ASC to form the AIM2 inflammasome. AIM2 inflammasome can induce

the maturation and release of inflammatory factors such as IL-1β and IL-18, triggering an inflammatory response (Wang et al., 2019). Studies over the past decades have shown that AIM2 inflammasome plays a vital role in many kinds of inflammatory diseases, immune diseases, and cancers (Man et al., 2016; Kumari et al., 2020; Zhu H. et al., 2021).

The assembly and activation of AIM2 inflammasome and subsequent inflammatory response was shown in **Figure 1**. The primary role of AIM2 is to initiate the assembly process of the inflammasome. ASC acts as an inflammasome adaptor protein, connecting upstream AIM2 and downstream Caspase-1. Caspase-1 is the effector protein of AIM2 inflammasome, and the activated caspase-1 leads to proteolytic cleavage of IL-1β and IL-18. The massive release of downstream inflammatory cytokines (IL-1β and IL-18) directly affects the host's innate immune regulation to infection and injury, induces acute and chronic inflammatory responses, and participates in the occurrence and development of these diseases, such as skin diseases, chronic kidney disease, cardiovascular diseases, neuronal diseases, and diabetes mellitus (Sharma et al., 2019).

3 GENETIC AND EPIGENETIC ASSOCIATIONS BETWEEN *AIM2* GENE AND PSORIASIS

3.1 Genetic Associations Between *AIM2* Gene and Psoriasis

In the early stage, sequencing results showed that the coding region of *AIM2* gene with a high frequency of frameshift and missense mutations in primary high-level microsatellite instability (MSI-H) colon cancers and cell lines (Woerner et al., 2007). In 2015, an exome-wide association study in large-scale individuals was performed to investigate the coding variants in psoriasis. *AIM2* gene was firstly identified as a susceptibility gene for psoriasis at the genome-wide level, this gene locates at an important topological position in the gene–gene interaction network. In addition, Zuo et al. (2015) predicted that a variant (rs2276405) in *AIM2* gene affects AIM2 protein structure (Glu32Lys). The chemical properties of Glu (the chemical nature of Glu residue is acidic) and Lys (the chemical nature of Lys residue is alkaline) are completely opposite, which may destabilize the alpha-helix motif. Researchers investigated the correlation between the genetic pattern of *AIM2* gene polymorphism and the psoriasis phenotype (Li et al., 2016). Genotype and allele distribution of *AIM2* gene showed that allele A was the minor allele and G was the risk allele. Genetic pattern analysis showed that the dominant pattern was the best genetic pattern for *AIM2* gene polymorphism loci. Compared with the controls, the distribution of the dominant inheritance pattern was higher in psoriasis area and severity index (PASI) score ≤20 than that of PASI score >20, and family history of negative patients is more significantly different statistically than positive family history.

3.2 Epigenetic Associations Between *AIM2* Gene and Psoriasis

Epigenetic associations between *AIM2* gene and diseases have also been studied. Hypermethylation of the *AIM2* promoter conferred insensitivity to IFN-γ-induced *AIM2* expression of MSI-H colon cancer cell lines, demonstrating the inactivation of *AIM2* was regulated by epigenetic factors (Woerner et al., 2007). In 2016, an epigenome association study (EWAS) of psoriasis was conducted in the Chinese Han population, and the results showed that three CpG sites (cg17217296, cg17515347, and cg07195224) in the promoter region of *AIM2* gene were significantly associated with psoriasis (Li et al., 2016). Assay for transposase-accessible chromatin using sequencing (ATAC-seq) was used to explore the landscape of chromatin accessibility of psoriatic skin tissue (PP) and nonpsoriatic skin tissue (PN) from patients with psoriasis and normal skin tissue (NN) from healthy individuals, the results show that *AIM2* gene promoter region was specifically more accessible in PP, and contained a CpG site (cg07195224), which was previously reported to be significantly hypomethylated in psoriasis (Tang L. et al., 2021). In addition, it was observed that the intensity of the promoter-associated peak of *AIM2* was negatively correlated with the methylation level of cg07195224 but positively correlated with *AIM2* mRNA expression level. Meanwhile, the methylation level of cg07195224 was strongly and negatively associated with *AIM2* mRNA expression level (Tang L. et al., 2021). Studies will be needed to explore the correlation between epigenetic regulation and *AIM2* gene expression in the future and to further clarify its regulating effect in the pathogenesis of psoriasis.

4 ROLES OF *AIM2* GENE AND AIM2 INFLAMMASOME IN THE PATHOGENESIS OF PSORIASIS

4.1 AIM2 Inflammasome Mediated Inflammatory Response Involved in Psoriasis

The roles of *AIM2* gene in the pathogenesis of psoriasis have been studied (**Table 1**). AIM2 inflammasome activity may represent a potential trigger for the occurrence and development of inflammatory diseases. Kopfnagel et al. (2011) found that the AIM2 inflammasome is active in human keratinocytes, triggering IL-1β secretion with an important role in inflammatory processes. In cultured keratinocytes, the expression of *AIM2* gene was induced by INF-γ; cytoplasmic DNA can trigger the release of IL-1β through the AIM2 inflammasome (Dombrowski et al., 2011). In $CD14^+$ and $CD16^+$ monocyte subsets in the blood of patients with psoriasis, the *AIM2* gene expression level was significantly higher than that of the control group, indicating that AIM2 inflammasome exists in the immune cell subsets in the peripheral blood of psoriasis patients in a stimulated state (Verma et al., 2021).

The expression of *AIM2* gene in lesional skin from psoriasis patients was explored in several studies (Dombrowski et al., 2011;

TABLE 1 | Roles of *AIM2* gene and AIM2 inflammasome in the pathogenesis of psoriasis.

Studies/Year	Objects	Main findings
Kopfnagel et al. (2011)	Keratinocyte	AIM2 inflammasome is active in human keratinocytes and triggers IL-1β secretion, which represents a potential trigger factor for the development and maintenance of inflammatory skin diseases
Dombrowski et al. (2011)	Skin tissue, Keratinocyte	Abundant cytoplasmic DNA and increased *AIM2* expression were detected in psoriatic skin lesions; *AIM2* expression was induced by INF-γ in cultured keratinocytes, and cytoplasmic DNA can trigger the release of IL-1β through the AIM2 inflammasome
de Koning et al. (2012)	Skin tissue	AIM2 protein expression is significantly upregulated in the psoriatic epidermis. *AIM2* expression was dynamics in human tissues and primary cells, restricted expression in Langerhans cell and melanocyte of the normal epidermis, but with a strong upregulation in subpopulations of epidermal keratinocytes under inflammatory conditions
Göblös et al. (2016)	Keratinocyte	Gene-specific silencing of *CARD18* in cells treated with poly (dA:dT) resulted in a significant decrease in *AIM2* gene expression and significantly reduced *Caspase-1* mRNA expression, which indicates that *CARD18* might indeed contribute to the fine-tuning of keratinocyte innate immune processes
Šahmatova et al. (2017)	Skin tissue	*AIM2* gene was expressed at an increased level in psoriatic skin
Yuan et al. (2020)	Skin tissue	High expression of *AIM2* gene can be detected in the epidermis of psoriatic skin lesions. Neutrophil extracellular trapping net (NET) may promote the expression of *AIM2* gene by activating keratinocytes. The secretion of IL-1β accelerates the inflammatory process of psoriasis
Verma et al. (2021)	Blood	An increased *AIM2* expression in the $CD14^+$ and $CD16^+$ subsets of patients was observed, suggesting that *AIM2* exists in a primed state in the immune cell subsets in the peripheral blood of the psoriasis patients
Tang H. et al. (2021)	Skin tissue, Keratinocyte	AURKA promotes the occurrence and development of psoriatic inflammation by blocking autophagy-mediated suppression of the AIM2 inflammasome
Zhao et al. (2022)	Skin tissue	The AIM2 total fluorescence intensity in $CD4^+$ Trm cells in patients with SCLE and localized DLE was higher than in patients with psoriasis. The expression of *AIM2* gene in skin $CD4^+$ Trm cells can be a significant indicator to distinguish patients with ACLE from those with localized DLE and SCLE.

de Koning et al., 2012; Šahmatova et al., 2017; Yuan et al., 2020). Compared with skin tissues from healthy individuals, abundant cytoplasmic DNA and increased *AIM2* gene expression were detected in psoriatic skin lesions (Dombrowski et al., 2011). Šahmatova et al. (2017) also identified that the *AIM2* gene was expressed at an increased level in psoriatic skin lesions. Yuan et al. (2020) found that the formation of neutrophil extracellular trapping net (NET) structure and the high expression of *AIM2* gene can be detected in the epidermis of skin lesions from psoriasis patients. NET may promote the expression of *AIM2* gene by activating keratinocytes, and the secretion of IL-1β accelerates the inflammatory process of psoriasis. In addition, the expression of *AIM2* gene is also significantly upregulated in several inflammatory skin disorders, including contact dermatitis, venous ulcers, atopic dermatitis, and experimental wounds (de Koning et al., 2012). The expression of *AIM2* gene was dynamic in human tissues and primary cells, with restricted expression in Langerhans cell and melanocytes of the normal epidermis, but a strong upregulation in subpopulations of epidermal keratinocytes under inflammatory conditions (de Koning et al., 2012).

The IL-23/Th17 axis plays a central role in the development of psoriasis (Hawkes et al., 2017), some biologics (IL-17 antagonists: Secukinumab, Ixekizumab, and Brodalumab; IL-23 antagonists: Risankizumab, Guselkumab, and Tildrakizumab) focus on this signaling pathway have been used for the treatment of moderate-severe psoriasis (Sharma et al., 2022). AIM2 induces AIM2-dependent release of IL-18 and IL-1β; furthermore, the active IL-1β and IL-18 involve in IL23/Th17 pathway and then induce many kinds of chemokines and inflammatory cytokines (Ciążyńska et al., 2021). Taken together, the aforementioned studies indicate that the activation of the AIM2 inflammasome may be a potential trigger for the development of psoriasis, and the inflammatory response mediated by the AIM2 inflammasome plays an important role in the initial onset and persistence of psoriasis.

4.2 Factors Regulate AIM2 Inflammasome Signaling Pathway in Psoriasis

Recent studies have found that certain factors regulate the AIM2 inflammasome signaling pathway and participate in the pathogenesis of psoriasis, including prokineticin 2 (PK2), caspase recruitment domain family member 18 (CARD18), aurora kinase A (AURKA), TLR-7, TLR-8, and TLR-9 antagonists.

4.2.1 PK2

PK2 is a psoriasis-specific factor that is highly expressed in mouse and human psoriatic skins, but is not significantly expressed in other autoimmune diseases, such as inflammatory bowel diseases, atherosclerosis, and diabetes; PK2 significantly increases the expression of Caspase-1, and also strongly up-regulates the AIM2 inflammasome signaling pathway in monocytic THP-1 cells, which is engaged by AIM2 to promote the synthesis and secretion of proinflammatory cytokine IL-1β (He et al., 2016). Thus, PK2 regulates the AIM2 inflammasome signaling pathway involved in the pathogenesis of psoriasis.

4.2.2 CARD18

CARD18 is highly expressed in psoriatic noninvolved epidermis compared to healthy skin epidermis (Szabó et al., 2014). CARD18 involves AIM2 inflammasome-mediated keratinocyte functions and modifies the inflammatory process in keratinocytes. Silencing of *CARD18* in keratinocyte cells treated with poly (dA:dT) resulted in a significant decrease in *AIM2* gene expression and significantly reduced *Caspase-1* mRNA expression (Göblös et al., 2016), which indicates that

CARD18 may contribute to the fine-tuning of innate immune processes.

4.2.3 AURKA

AURKA is a member of the serine/threonine kinases family, which plays a critical role in the suppression of autophagy (Zhang S. et al., 2017), and is elevated in lesional psoriatic tissue (Liu et al., 2011). A recent study found that AURKA promotes the occurrence and development of psoriatic inflammation by blocking autophagy-mediated suppression of the AIM2 inflammasome (Tang H. et al., 2021).

4.2.4 TLR-7, TLR-8, and TLR-9 Antagonists

Immune modulatory oligonucleotides and small molecular weight compounds, IMO-3100, IMO-8400, and IMO-9200 target TLR-7, TLR-8, and TLR-9, respectively. These are under clinical investigation for their effectiveness in the treatment of psoriasis. Chemical compounds, such as AS-2444697, PF-05387252, PF-05388169, PF-06650833, ML120B, and PHA-408, can inhibit TLR signaling (Gao et al., 2017). Jiang et al. (2013) evaluated an antagonist of TLR-7, TLR-8, and TLR-9 as a therapeutic agent in an IL-23-induced psoriasis model in C57BL/6 mice. Treatment with an antagonist reduced the expression of inflammasome components (including NLRP3, AIM2, and antimicrobial peptides) in the dermis, which indicated that targeting TLR-7, TLR-8, and TLR-9 may provide a method for neutralizing the multiple inflammatory pathways that are involved in psoriasis.

4.3 How Fra-1 Regulates AIM2 and What's the Regulatory Effect in Psoriasis?

Recently, bioinformatics prediction and luciferase reporter assay showed that Fra-1 targeted binding to 363 bp and 57 bp upstream of the *AIM2* gene transcription start site (Tang L et al., 2021). Fra-1 is encoded by the *FOSL1* gene and plays a role in cell proliferation and differentiation, gene expression and regulation, and the occurrence and progression of psoriasis (Talotta et al., 2020). *FOSL1* knockdown inhibited IL-22-induced proliferation and enhanced keratinocyte apoptosis, whereas IL-22 stimulation and *FOSL1* overexpression further enhanced keratinocyte proliferation (Meng et al., 2021). The expression of *FOSL1* was significantly increased in lesional psoriatic skin and positively correlated with the PASI score (Sobolev et al., 2011). The aforementioned findings show that the high expression of *FOSL1* in lesional psoriatic skin is one of the markers of the pathological activity of psoriasis and that Fra-1 plays an important role in the pathogenesis of psoriasis. However, Fra-1 needs to be further investigated on how it regulates the expression of *AIM2* in psoriasis, which is still unclear.

4.4 The Expression of *AIM2* Gene Can Be a Biomarker to Distinguish Different Subtypes of Psoriasis?

The expression level of *AIM2* gene in $CD4^+$ tissue-resident memory T ($CD4^+$ Trm) cells was measured in the patients with acute cutaneous lupus erythematosus (ACLE), subacute CLE (SCLE), localized discoid lupus erythematosus (localized DLE), psoriasis, and other inflammatory skin diseases (Zhao et al., 2022). The results showed that *AIM2* expression in $CD4^+$ Trm cells was significantly lower in patients with ACLE than in localized DLE and SCLE. In psoriasis patients, $CD4^+$ Trm cells were mainly located in the epidermis of skin lesions. The AIM2 total fluorescence intensity in $CD4^+$ Trm cells in patients with SCLE and localized DLE were higher than in patients with psoriasis. Compared to ACLE with localized DLE and/or SCLE, the receiver operating characteristic curve for *AIM2* expression in $CD4^+$ Trm cells had different sensitivities and specificities at different cutoff values. Therefore, *AIM2* expression in skin $CD4^+$ Trm cells can be a significant indicator of distinguishing patients with ACLE from those patients with SCLE and localized DLE. Psoriasis is also divided into subtypes with different clinical manifestations, including psoriasis vulgaris, erythrodermic psoriasis, pustular psoriasis, and psoriatic arthritis (Zhu C. et al., 2021). It is worthy to further investigate whether *AIM2* expression can be a biomarker to distinguish different subtypes of psoriasis. A valuable study will be carried out in the future to explore this scientific question.

5 AIM2 INFLAMMASOME IS A PROMISING THERAPEUTIC TARGET FOR PSORIASIS

5.1 AIM2 Correlates With Therapeutic Efficacy of Psoriasis

Recent studies have shown that the expression of *AIM2* gene can be a biomarker to predict the benefit of therapy in patients with epithelial ovarian cancer (Hsu et al., 2021), melanoma (Fukuda et al., 2021), systemic lupus erythematosus (Yang et al., 2021) and heart failure (Onódi et al., 2021). These studies implicate its potential utility in predicting clinical treatment outcomes. Gong et al. calculated innate immune cell proportion in psoriatic skin by utilizing microarray data, results show that *AIM2* gene was negatively associated with resting mast cells but positively associated with activated dendritic cells (Gong and Wang, 2021). Brodalumab and Ustekinumab are used to treat moderate to severe plaque psoriasis, and most patients can get better treatment effects. Interestingly, *AIM2* gene was positively associated with the therapeutic efficacy of Brodalumab and negatively associated with Ustekinumab treatment response (Gong and Wang, 2021), which provide novel clues for clinical decisions on treatment for psoriasis. In addition, KEGG analysis shows pathways involving RIG-I-like receptor signaling, NOD-like receptor signaling, toll-like receptor signaling, and cytosolic DNA-sensing pathways, these pathways were significantly enriched and positively correlated with *AIM2* gene (Gong and Wang, 2021).

5.2 LL-37 Neutralizes Inflammation Mediated by AIM2 Inflammasome

The antimicrobial peptide LL-37 (also known as cathelicidin antimicrobial peptide, CAMP) is overexpressed in psoriatic lesions and acts as the critical factor that mediates plasmacytoid dendritic cells

(pDCs) activation in psoriasis (Lande et al., 2007). AIM2 inflammasomes activate pro-inflammatory processes and trigger IL-1β secretion in psoriatic lesions. However, IL-1β secretion was completely abolished when LL-37 and DNA were delivered together into keratinocytes. This may suggest that LL-37 can translocate into the cytosol of psoriatic keratinocytes and specifically neutralizes cytosolic DNA, thus acting as an inhibitor of AIM2 inflammasome activation (Dombrowski and Schauber, 2012). Topical treatment with vitamin D analogs decreases inflammation and pro-inflammatory cytokines, but strongly increases cathelicidin's expression (Lebwohl et al., 2004; Peric et al., 2009). However, repeated treatments with narrowband-UVB (NB-UVB) decreased skin inflammation in patients with psoriasis but increased vitamin D serum levels and the expression of cutaneous cathelicidin (Vähävihu et al., 2010). Thus, established therapies targeting the vitamin D pathway reduce inflammatory responses while increasing epidermal antimicrobial peptide expression in psoriatic lesions, the anti-inflammatory effect of LL-37 on the AIM2 inflammasome pathway may account for these observed effects (Reinholz et al., 2012). LL-37 can modulate the pathogenic response to nucleic acids and be helpful for the development of anti-inflammatory therapies.

5.3 Potential AIM2 Inflammasome Inhibitors for Psoriasis Treatment

At present, there are some AIM2 inflammasome inhibitors that have been explored to treat inflammatory diseases, including epigallocatechin gallate (EGCG), EFLA-945, obovatol, withaferin A (WFA), and RGFP966.

5.3.1 EGCG

EGCG is the most abundant main polyphenol component of green tea, and the alloy moiety of catechins possesses the most biological activities, including angiogenesis and anti-inflammatory effects (Kondo et al., 2002). The structure of EGCG and related information are shown in **Supplementary Figure S1**. EGCG inhibits the transfection of NF-κB and AP-1 to downregulate the expression of reactive oxygen species (ROS) and reactive nitrogen species (RNS) and decreases the production of inflammatory factors (Nagai et al., 2002). Studies have shown that EGCG attenuates AIM2-induced IL-1β secretion by inhibiting inflammasome IFN-γ secretion and dA:dT-induced ASC oligomerization in neonatal human epidermal keratinocytes (Yun et al., 2015).

5.3.2 EFLA-945

EFLA-945 is an extract from red grapevine leaf in traditional medicine in Japan (Rabe et al., 2011) and over-the-counter drugs in Europe and Japan (Hoshino et al., 2018) for its anti-inflammatory properties and circulatory benefits. Resveratrol (**Supplementary Figure S2**) and peonidin 3-O-glucoside (**Supplementary Figure S3**) are the major phytochemicals of EFLA 945. Study results showed that EFLA-945 could limit the entry of DNA into THP-1-derived macrophages, thereby inhibiting cytoplasmic DNA-dependent ASC and activation of the AIM2 inflammasome. Resveratrol and peonidin 3-O-glucoside are two major phytochemicals of EFL-945 that mediate this inhibition (Chung et al., 2020). Furthermore, EFLA-945 attenuated the associated pro-inflammatory response in localized skin lesions of an imiquimod-induced psoriasis-like mouse model, indicating that EFLA-945 may be beneficial for the treatment of psoriasis.

5.3.3 Obovatol

Obovatol (**Supplementary Figure S4**), a bisphenol chemical originating from Magnolia obovata, affects the AIM2 inflammasome by inhibiting the formation of ASC pyroptosome and the generation of mitochondrial ROS. Furthermore, obovatol has been used as a traditional treatment for inflammatory diseases (Kim et al., 2019) and neuroinflammation (Ock et al., 2010). Furthermore, obovatol disrupted inflammasome activation's initiation step, inhibited the transcription of inflammatory cytokines, and decreased serum IL-1β elevation in response to sodium urate crystals in mice (Kim et al., 2019). Whether obovatol has a therapeutic effect on psoriasis will be further studied.

5.3.4 WFA

WFA is an extract from the medicinal plant *Withania somnifera* and has various biological activities, including acting as an anti-inflammatory, angiogenesis, and anticancer (Mohan et al., 2004). The structure of EGCG and related information are shown in **Supplementary Figure S5**. Studies show that WFA regulates AIM2 inflammasome and Caspase-1 in THP-1 polarized macrophages; however, WFA treatment of M2 macrophages inhibits TGF-β compared with M1 secretion (Ngoungoure and Owona, 2019). Currently, further research is needed to determine whether WFA can affect the occurrence and progression of psoriasis by inhabiting the AIM2 inflammasome pathway.

5.3.5 RGFP966

Histone deacetylases 3 (HDAC3) modulates the acetylation of histone and non-histone proteins. RGFP966 (**Supplementary Figure S6**) is a selective inhibitor of HDAC3 (Zhang et al., 2020). Specifically, RGFP966 regulates the inflammatory process in stroke (Chen et al., 2012) and brain damage (Zhang et al., 2020). Lipopolysaccharide (LPS) stimulation caused time-dependent increases of HDAC3 and AIM2 inflammasome in primary cultured microglia. *AIM2* gene was spatiotemporally regulated by RGFP966, which was confirmed in an experimental mouse stroke model. RGFP966 can enhance STAT1 acetylation and decrease STAT1 phosphorylation, which may partially explain the negative regulatory effect of *AIM2* gene by RGFP966 (Zhang et al., 2020). This study indicated that RGFP966 alleviated the inflammatory process by regulating the AIM2 inflammasome. The further therapeutic effects of RGFP966 on psoriasis by regulating the AIM2 inflammasome will be investigated in the future.

6 THE DISADVANTAGES OF TARGETING *AIM2* GENE AS A POTENTIAL THERAPEUTIC TARGET

AIM2 gene act as a double-edged sword in the pathogenesis of some autoimmune diseases and cancers. The increased and decreased expression of *AIM2* gene with different roles in the development

of different diseases. *AIM2* gene plays the tumor-suppressive role in HPV-infected cervical cancer (So et al., 2018), breast cancer (Yoon et al., 2015), squamous cell carcinoma (Farshchian et al., 2017); while *AIM2* gene is regarded as a protective factor in melanoma (DeYoung et al., 1997; de Koning et al., 2014) and colorectal cancer (Wilson et al., 2015; Zhang Z. et al., 2017). Intriguingly, *AIM2* gene with two contrasting roles in different models or different disease stages of hepatocellular carcinoma (Martínez-Cardona et al., 2018; Shi et al., 2019). *AIM2* gene as a therapeutic target may with different effects on different diseases, therefore we should consider the advantages and disadvantages of targeting *AIM2* gene. For instance, when we use AIM2 inflammasome inhibitors or drugs to treat psoriatic patients also with cancer, in the future, we need to pay attention to its impact on cancer. Taken together, the advantages and disadvantages of targeting *AIM2* gene should be considered at the same time.

7 CONCLUSION AND PERSPECTIVES

In this review, we provided an overview of the research progress on the effect and role of *AIM2* gene and AIM2 inflammasome in psoriasis. We hope this provides further potential therapeutic targets for psoriasis. Although some potential inhibitors of the AIM2 inflammasome were explored, no AIM2 inflammasome target drug has been used in the clinical treatment of psoriasis. Therefore, further researches need to focus on the development of new drugs for psoriasis.

AUTHOR CONTRIBUTIONS

CS, BY, and YZ designed this study; JW, CS, JG, CH, WZ, CC, and XS collected and summarized the data from the literature search; JW, JG, and CS prepared the manuscript; JW, JG, CS, SJ, RK, CH, CC, WZ, YZ, and BY revised this manuscript. All authors contributed to the article and approved the submitted version.

FUNDINGS

This study was supported by grants from the Research Foundation of Peking University Shenzhen Hospital (No: JCYJ2020004), Shenzhen Sanming Project (No: SZSM201812059), Shenzhen Key Medical Discipline Construction Fund (No: SZXK040), Guangdong Basic and Applied Basic Research Foundation (No: 2021A1515110097), National Natural Science Foundation of China (No: 82103726), Natural Science Foundation of Anhui Province (No: 2008085QH427), and Research Fund of Anhui Institute of Translational Medicine (No: 2021ZHYX-C48).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.929162/full#supplementary-material

REFERENCES

Branisteanu, D., Cojocaru, C., Diaconu, R., Porumb, E., Alexa, A., Nicolescu, A., et al. (2022). Update on the Etiopathogenesis of Psoriasis (Review). *Exp. Ther. Med.* 23, 201. doi:10.3892/etm.2022.11124

Chen, Y.T., Zang, X.F., Pan, J., Zhu, X.L., Chen, F., Chen, Z.B., et al. (2012). Expression Patterns of Histone Deacetylases in Experimental Stroke and Potential Targets for Neuroprotection. *Clin. Exp. Pharmacol. Physiol.* 39, 751–758. doi:10.1111/j.1440-1681.2012.05729.x

Choubey, D., Walter, S., Geng, Y., and Xin, H. (2000). Cytoplasmic Localization of the Interferon-Inducible Protein that Is Encoded by theAIM2(absent in Melanoma) Gene from the 200-gene Family. *FEBS Lett.* 474, 38–42. doi:10.1016/s0014-5793(00)01571-4

Chung, I.C., Yuan, S.N., OuYang, C.N., Hu, S.I., Lin, H.C., Huang, K.Y., et al. (2020). EFLA 945 Restricts AIM2 Inflammasome Activation by Preventing DNA Entry for Psoriasis Treatment. *Cytokine* 127, 154951. doi:10.1016/j.cyto.2019.154951

Ciążyńska, M., Olejniczak-Staruch, I., Sobolewska-Sztychny, D., Narbutt, J., Skibińska, M., and Lesiak, A. (2021). The Role of NLRP1, NLRP3, and AIM2 Inflammasomes in Psoriasis: Review. *Int. J. Mol. Sci.* 22, 5898. doi:10.3390/ijms22115898

de Koning, H. D., Bergboer, J. G. M., van den Bogaard, E. H., van Vlijmen-Willems, I. M. J. J., Rodijk-Olthuis, D., Simon, A., et al. (2012). Strong Induction of AIM2 Expression in Human Epidermis in Acute and Chronic Inflammatory Skin Conditions. *Exp. Dermatol.* 21, 961–964. doi:10.1111/exd.12037

de Koning, H. D., van Vlijmen-Willems, I. M. J. J., Zeeuwen, P. L. J. M., Blokx, W. A. M., and Schalkwijk, J. (2014). Absent in Melanoma 2 Is Predominantly Present in Primary Melanoma and Primary Squamous Cell Carcinoma, but Largely Absent in Metastases of Both Tumors. *J. Am. Acad. Dermatol.* 71, 1012–1015. doi:10.1016/j.jaad.2014.06.012

DeYoung, K. L., Ray, M. E., Su, Y. A., Anzick, S. L., Johnstone, R. W., Trapani, J. A., et al. (1997). Cloning a Novel Member of the Human Interferon-Inducible Gene Family Associated with Control of Tumorigenicity in a Model of Human Melanoma. *Oncogene* 15, 453–457. doi:10.1038/sj.onc.1201206

Dombrowski, Y., Peric, M., Koglin, S., Kammerbauer, C., Göß, C., Anz, D., et al. (2011). Cytosolic DNA Triggers Inflammasome Activation in Keratinocytes in Psoriatic Lesions. *Sci. Transl. Med.* 3, 82ra38. doi:10.1126/scitranslmed.3002001

Dombrowski, Y., and Schauber, J. (2012). Cathelicidin LL-37: a Defense Molecule with a Potential Role in Psoriasis Pathogenesis. *Exp. Dermatol.* 21, 327–330. doi:10.1111/j.1600-0625.2012.01459.x

Fagerberg, L., Hallström, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., et al. (2014). Analysis of the Human Tissue-specific Expression by Genome-wide Integration of Transcriptomics and Antibody-Based Proteomics. *Mol. Cell. Proteomics* 13, 397–406. doi:10.1074/mcp.M113.035600

Farshchian, M., Nissinen, L., Siljamäki, E., Riihilä, P., Piipponen, M., Kivisaari, A., et al. (2017). Tumor Cell-specific AIM2 Regulates Growth and Invasion of Cutaneous Squamous Cell Carcinoma. *Oncotarget* 8, 45825–45836. doi:10.18632/oncotarget.17573

Fernandes-Alnemri, T., Yu, J. W., Datta, P., Wu, J., and Alnemri, E. S. (2009). AIM2 Activates the Inflammasome and Cell Death in Response to Cytoplasmic DNA. *Nature* 458, 509–513. doi:10.1038/nature07710

Fukuda, K., Okamura, K., Riding, R. L., Fan, X., Afshari, K., Haddadi, N.-S., et al. (2021). AIM2 Regulates Anti-tumor Immunity and Is a Viable Therapeutic Target for Melanoma. *J. Exp. Med.* 218, e20200962. doi:10.1084/jem.20200962

Gao, J., Shen, X., Ko, R., Huang, C., and Shen, C. (2021). Cognitive Process of Psoriasis and its Comorbidities: From Epidemiology to Genetics. *Front. Genet.* 12, 735124. doi:10.3389/fgene.2021.735124

Gao, W., Xiong, Y., Li, Q., and Yang, H. (2017). Inhibition of Toll-like Receptor Signaling as a Promising Therapy for Inflammatory Diseases: A Journey from Molecular to Nano Therapeutics. *Front. Physiol.* 8, 508. doi:10.3389/fphys.2017.00508

Göblös, A., Danis, J., Vas, K., Bata-Csörgő, Z., Kemény, L., and Széll, M. (2016). Keratinocytes Express Functional CARD18, a Negative Regulator of Inflammasome Activation, and its Altered Expression in Psoriasis May Contribute to Disease Pathogenesis. *Mol. Immunol.* 73, 10–18. doi:10.1016/j.molimm.2016.03.009

Gong, X., and Wang, W. (2021). Profiles of Innate Immune Cell Infiltration and Related Core Genes in Psoriasis. *BioMed Res. Int.* 2021, 6656622. doi:10.1155/2021/6656622

Hawkes, J. E., Chan, T. C., and Krueger, J. G. (2017). Psoriasis Pathogenesis and the Development of Novel Targeted Immune Therapies. *J. Allergy Clin. Immunol.* 140, 645–653. doi:10.1016/j.jaci.2017.07.004

He, X., Shen, C., Lu, Q., Li, J., Wei, Y., He, L., et al. (2016). Prokineticin 2 Plays a Pivotal Role in Psoriasis. *EBioMedicine* 13, 248–261. doi:10.1016/j.ebiom.2016.10.022

Hornung, V., Ablasser, A., Charrel-Dennis, M., Bauernfeind, F., Horvath, G., Caffrey, D. R., et al. (2009). AIM2 Recognizes Cytosolic dsDNA and Forms a Caspase-1-Activating Inflammasome with ASC. *Nature* 458, 514–518. doi:10.1038/nature07725

Hoshino, T., Muto, N., Tsukada, S., Nakamura, T., and Maegawa, H. (2018). European Ethnopharmaceuticals for Self-Medication in Japan: Review Experience of Vitis vinifera L., Folium Extract and Vitex Agnus-Castus L., Fructus Extract as OTC Drugs. *Medicines* 5, 3. doi:10.3390/medicines5010003

Hsu, P.C., Chao, T.K., Chou, Y.C., Yu, M.H., Wang, Y.C., Lin, Y.H., et al. (2021). AIM2 Inflammasome in Tumor Cells as a Biomarker for Predicting the Treatment Response to Antiangiogenic Therapy in Epithelial Ovarian Cancer Patients. *J. Clin. Med.* 10, 4529. doi:10.3390/jcm10194529

Jiang, W., Zhu, F.G., Bhagat, L., Yu, D., Tang, J. X., Kandimalla, E. R., et al. (2013). A Toll-like Receptor 7, 8, and 9 Antagonist Inhibits Th1 and Th17 Responses and Inflammasome Activation in a Model of IL-23-induced Psoriasis. *J. Invest. Dermatol.* 133, 1777–1784. doi:10.1038/jid.2013.57

Kim, J., Ahn, H., Han, B. C., Shin, H., Kim, J.C., Jung, E.M., et al. (2019). Obovatol Inhibits NLRP3, AIM2, and Non-canonical Inflammasome Activation. *Phytomedicine* 63, 153019. doi:10.1016/j.phymed.2019.153019

Kondo, T., Ohta, T., Igura, K., Hara, Y., and Kaji, K. (2002). Tea Catechins Inhibit Angiogenesis *In Vitro*, Measured by Human Endothelial Cell Growth, Migration and Tube Formation, through Inhibition of VEGF Receptor Binding. *Cancer Lett.* 180, 139–144. doi:10.1016/S0304-3835(02)00007-1

Kopfnagel, V., Wittmann, M., and Werfel, T. (2011). Human Keratinocytes Express AIM2 and Respond to dsDNA with IL-1β Secretion. *Exp. Dermatol.* 20, 1027–1029. doi:10.1111/j.1600-0625.2011.01382.x

Kumari, P., Russo, A. J., Shivcharan, S., and Rathinam, V. A. (2020). AIM2 in Health and Disease: Inflammasome and beyond. *Immunol. Rev.* 297, 83–95. doi:10.1111/imr.12903

Lande, R., Gregorio, J., Facchinetti, V., Chatterjee, B., Wang, Y.H., Homey, B., et al. (2007). Plasmacytoid Dendritic Cells Sense Self-DNA Coupled with Antimicrobial Peptide. *Nature* 449, 564–569. doi:10.1038/nature06116

Lebwohl, M., Menter, A., Koo, J., and Feldman, S. R. (2004). Combination Therapy to Treat Moderate to Severe Psoriasis. *J. Am. Acad. Dermatol.* 50, 416–430. doi:10.1016/j.jaad.2002.12.002

Li, G., Tao, L., Xu, Y., Dou, J., Wang, W., Zheng, X., et al. (2016). Genetic Model Analysis of AIM2 Single Nucleotide Polymorphism in Han Chinese Patients with Psoriasis Vulgaris. *Acta Univ. Med. Anhui* 51, 1486-1490. doi:10.19405/j.cnki.issn1000-1492.2016.10.020

Liang, D. W., Wang, L. Y., and Zhu, X. F. (2021). Research Advances in Correlation between Absent in Melanoma 2 and Psoriasis. *J. Clin. Med. Pract.* 25, 122-124. doi:10.7619/jcmp.20200520

Liu, Y., Luo, W., and Chen, S. (2011). Comparison of Gene Expression Profiles Reveals Aberrant Expression of FOXO1, Aurora A/B and EZH2 in Lesional Psoriatic Skins. *Mol. Biol. Rep.* 38, 4219–4224. doi:10.1007/s11033-010-0544-x

Man, S. M., Karki, R., and Kanneganti, T.D. (2016). AIM2 Inflammasome in Infection, Cancer, and Autoimmunity: Role in DNA Sensing, Inflammation, and Innate Immunity. *Eur. J. Immunol.* 46, 269–280. doi:10.1002/eji.201545839

Martínez-Cardona, C., Lozano-Ruiz, B., Bachiller, V., Peiró, G., Algaba-Chueca, F., Gómez-Hurtado, I., et al. (2018). AIM2 Deficiency Reduces the Development of Hepatocellular Carcinoma in Mice. *Int. J. Cancer* 143, 2997–3007. doi:10.1002/ijc.31827

Meng, J., Chen, F.R., Yan, W.J., and Lin, Y.K. (2021). MiR-15a-5p Targets FOSL1 to Inhibit Proliferation and Promote Apoptosis of Keratinocytes via MAPK/ERK Pathway. *J. Tissue Viability* 30, 544–551. doi:10.1016/j.jtv.2021.08.006

Mohan, R., Hammers, H., Bargagna-mohan, P., Zhan, X., Herbstritt, C., Ruiz, A., et al. (2004). Withaferin A Is a Potent Inhibitor of Angiogenesis. *Angiogenesis* 7, 115–122. doi:10.1007/s10456-004-1026-3

Nagai, K., Jiang, M. H., Hada, J., Nagata, T., Yajima, Y., Yamamoto, S., et al. (2002). (−)-Epigallocatechin Gallate Protects against NO Stress-Induced Neuronal Damage after Ischemia by Acting as an Anti-oxidant. *Brain Res.* 956, 319–322. doi:10.1016/S0006-8993(02)03564-3

Nedoszytko, B., Szczerkowska-Dobosz, A., Stawczyk-Macieja, M., Owczarczyk-Saczonek, A., Reich, A., Bartosińska, J., et al. (2020). Pathogenesis of Psoriasis in the "omic" Era. Part II. Genetic, Genomic and Epigenetic Changes in Psoriasis. *Postepy Dermatol. Alergol.* 37, 283–298. doi:10.5114/ada.2020.96243

Ngoungoure, F. P., and Owona, B. A. (2019). Withaferin A Modulates AIM2 Inflammasome and Caspase-1 Expression in THP-1 Polarized Macrophages. *Exp. Cell Res.* 383, 111564. doi:10.1016/j.yexcr.2019.111564

Ock, J., Han, H. S., Hong, S. H., Lee, S. Y., Han, Y.-M., Kwon, B.M., et al. (2010). Obovatol Attenuates Microglia-Mediated Neuroinflammation by Modulating Redox Regulation. *Br. J. Pharmacol.* 159, 1646–1662. doi:10.1111/j.1476-5381.2010.00659.x

Onódi, Z., Ruppert, M., Kucsera, D., Sayour, A. A., Tóth, V. E., Koncsos, G., et al. (2021). AIM2-driven Inflammasome Activation in Heart Failure. *Cardiovasc. Res.* 117, 2639–2651. doi:10.1093/cvr/cvab202

Peric, M., Koglin, S., Dombrowski, Y., Gross, K., Bradac, E., Büchau, A., et al. (2009). Vitamin D Analogs Differentially Control Antimicrobial Peptide/"Alarmin" Expression in Psoriasis. *PloS One* 4, e6340. doi:10.1371/journal.pone.0006340

Qi, M., Dai, D., Liu, J., Li, Z., Liang, P., Wang, Y., et al. (2020). AIM2 Promotes the Development of Non-small Cell Lung Cancer by Modulating Mitochondrial Dynamics. *Oncogene* 39, 2707–2723. doi:10.1038/s41388-020-1176-9

Rabe, E., Stücker, M., Esperester, A., Schäfer, E., and Ottillinger, B. (2011). Efficacy and Tolerability of a Red-Vine-Leaf Extract in Patients Suffering from Chronic Venous Insufficiency - Results of a Double-Blind Placebo-Controlled Study. *Eur. J. Vasc. Endovascular Surg.* 41, 540–547. doi:10.1016/j.ejvs.2010.12.003

Reinholz, M., Ruzicka, T., and Schauber, J. (2012). Cathelicidin LL-37: an Antimicrobial Peptide with a Role in Inflammatory Skin Disease. *Ann. Dermatol.* 24, 126–135. doi:10.5021/ad.2012.24.2.126

Šahmatova, L., Sügis, E., Šunina, M., Hermann, H., Prans, E., Pihlap, M., et al. (2017). Signs of Innate Immune Activation and Premature Immunosenescence in Psoriasis Patients. *Sci. Rep.* 7, 7553. doi:10.1038/s41598-017-07975-2

Samotij, D., Nedoszytko, B., Bartosińska, J., Batycka-Baran, A., Czajkowski, R., Dobrucki, I., et al. (2020). Pathogenesis of Psoriasis in the "omic" Era. Part I. Epidemiology, Clinical Manifestation, Immunological and Neuroendocrine Disturbances. *Postepy Dermatol. Alergol.* 37, 135–153. doi:10.5114/ada.2020.94832

Sharma, A., Upadhyay, D. K., Gupta, G. D., Narang, R. K., and Rai, V. K. (2022). IL-23/Th17 Axis: A Potential Therapeutic Target of Psoriasis. *Curr. Drug. Res. Rev.* 14, 24–36. doi:10.2174/2589977513666210707114520

Sharma, B. R., Karki, R., and Kanneganti, T. D. (2019). Role of AIM2 Inflammasome in Inflammatory Diseases, Cancer and Infection. *Eur. J. Immunol.* 49, 1998–2011. doi:10.1002/eji.201848070

Shi, X., Wang, L., Ren, L., Li, J., Li, S., Cui, Q., et al. (2019). Dihydroartemisinin, an Antimalarial Drug, Induces Absent in Melanoma 2 Inflammasome Activation and Autophagy in Human Hepatocellular Carcinoma HepG2215 Cells. *Phytotherapy Res.* 33, 1413–1425. doi:10.1002/ptr.6332

Shin, J. I., Lee, K. H., Joo, Y. H., Lee, J. M., Jeon, J., Jung, H. J., et al. (2019). Inflammasomes and Autoimmune and Rheumatic Diseases: A Comprehensive Review. *J. Autoimmun.* 103, 102299. doi:10.1016/j.jaut.2019.06.010

So, D., Shin, H. W., Kim, J., Lee, M., Myeong, J., Chun, Y. S., et al. (2018). Cervical Cancer Is Addicted to SIRT1 Disarming the AIM2 Antiviral Defense. *Oncogene* 37, 5191–5204. doi:10.1038/s41388-018-0339-4

Sobolev, V. V., Zolotorenko, A. D., Soboleva, A. G., Elkin, A. M., Il'ina, S. A., Serov, D. N., et al. (2011). Effects of Expression of Transcriptional Factor AP-1 FOSL1 Gene on Psoriatic Process. *Bull. Exp. Biol. Med.* 150, 632–634. doi:10.1007/s10517-011-1208-0

Szabó, K., Bata-Csörgő, Z., Dallos, A., Bebes, A., Francziszti, L., Dobozy, A., et al. (2014). Regulatory Networks Contributing to Psoriasis Susceptibility. *Acta Derm. Venerol.* 94, 380–385. doi:10.2340/00015555-1708

Talotta, F., Casalino, L., and Verde, P. (2020). The Nuclear Oncoprotein Fra-1: a Transcription Factor Knocking on Therapeutic Applications' Door. *Oncogene* 39, 4491–4506. doi:10.1038/s41388-020-1306-4

Tang, H., Tang, X., Guo, Z., Cheng, H., Zheng, X., Chen, G., et al. (2021). AURKA Facilitates the Psoriasis-Related Inflammation by Impeding Autophagy-Mediated AIM2 Inflammasome Suppression. *Immunol. Lett.* 240, 98–105. doi:10.1016/j.imlet.2021.10.004

Tang, L., Wang, M., Shen, C., Wen, L., Li, M., Wang, D., et al. (2021). Assay for Transposase-Accessible Chromatin Using Sequencing Analysis Reveals a Widespread Increase in Chromatin Accessibility in Psoriasis. *J. Invest. Dermatol.* 141, 1745–1753. doi:10.1016/j.jid.2020.12.031

Vähävihu, K., Ala-Houhala, M., Peric, M., Karisola, P., Kautiainen, H., Hasan, T., et al. (2010). Narrowband Ultraviolet B Treatment Improves Vitamin D Balance and Alters Antimicrobial Peptide Expression in Skin Lesions of Psoriasis and Atopic Dermatitis. *Br. J. Dermatol.* 163, 321–328. doi:10.1111/j.1365-2133.2010.09767.x

Vakrakou, A. G., Svolaki, I. P., Evangelou, K., Gorgoulis, V. G., and Manoussakis, M. N. (2020). Cell-autonomous Epithelial Activation of AIM2 (Absent in Melanoma-2) Inflammasome by Cytoplasmic DNA Accumulations in Primary Sjögren's Syndrome. *J. Autoimmun.* 108, 102381. doi:10.1016/j.jaut.2019.102381

Verma, D., Fekri, S. Z., Sigurdardottir, G., Bivik Eding, C., Sandin, C., and Enerbäck, C. (2021). Enhanced Inflammasome Activity in Patients with Psoriasis Promotes Systemic Inflammation. *J. Invest. Dermatol.* 141, 586–595. doi:10.1016/j.jid.2020.07.012

Wang, B., Tian, Y., and Yin, Q. (2019). AIM2 Inflammasome Assembly and Signaling. *Adv. Exp. Med. Biol.* 1172, 143–155. doi:10.1007/978-981-13-9367-9_7

Wilson, J. E., Petrucelli, A. S., Chen, L., Koblansky, A. A., Truax, A. D., Oyama, Y., et al. (2015). Inflammasome-independent Role of AIM2 in Suppressing Colon Tumorigenesis via DNA-PK and Akt. *Nat. Med.* 21, 906–913. doi:10.1038/nm.3908

Woerner, S. M., Kloor, M., Schwitalle, Y., Youmans, H., Doeberitz, M. v. K., Gebert, J., et al. (2007). The Putative Tumor suppressorAIM2is Frequently Affected by Different Genetic Alterations in Microsatellite Unstable Colon Cancers. *Genes Chromosom. Cancer* 46, 1080–1089. doi:10.1002/gcc.20493

Yang, M., Long, D., Hu, L., Zhao, Z., Li, Q., Guo, Y., et al. (2021). AIM2 Deficiency in B Cells Ameliorates Systemic Lupus Erythematosus by Regulating Blimp-1-Bcl-6 Axis-Mediated B-Cell Differentiation. *Sig Transduct. Target Ther.* 6, 341. doi:10.1038/s41392-021-00725-x

Yoon, N., Park, M. S., Peltier, G. C., and Lee, R. H. (2015). Pre-activated Human Mesenchymal Stromal Cells in Combination with Doxorubicin Synergistically Enhance Tumor-Suppressive Activity in Mice. *Cytotherapy* 17, 1332–1341. doi:10.1016/j.jcyt.2015.06.009

Yuan, X., Shao, S., and Wang, G. (2020). Neutrophil Extracellular Traps Contribute to the Occurrence and Development of Psoriasis via Activating AIM2 Inflammasomes in Keratinocytes. *Chin. J. Dermatol.* 53, 324-329. doi:10.35541/cjd.20190897

Yun, M., Seo, G., Lee, J. Y., Chae, G. T., and Lee, S. B. (2015). Epigallocatechin-3-gallate Attenuates the AIM2-Induced Secretion of IL-1β in Human Epidermal Keratinocytes. *Biochem. Biophysical Res. Commun.* 467, 723–729. doi:10.1016/j.bbrc.2015.10.075

Zhang, M. J., Zhao, Q. C., Xia, M. X., Chen, J., Chen, Y. T., Cao, X., et al. (2020). The HDAC3 Inhibitor RGFP966 Ameliorated Ischemic Brain Damage by Downregulating the AIM2 Inflammasome. *FASEB J.* 34, 648–662. doi:10.1096/fj.201900394RRR

Zhang, S., Li, J., Zhou, G., Mu, D., Yan, J., Xing, J., et al. (2017). Aurora-A Regulates Autophagy through the Akt Pathway in Human Prostate Cancer. *Cancer Biomark* 19, 27–34. doi:10.3233/cbm-160238

Zhang, Z., Dong, X., Yang, X., Wan, D., Sun, L., Gu, M., et al. (2017). Expression and Clinical Significance of Absent in Melanoma 2 in Colorectal Cancer. *Biomed. Pharmacother.* 94, 843–849. doi:10.1016/j.biopha.2017.07.161

Zhao, Z., Zhu, H., Li, Q., Liao, W., Chen, K., Yang, M., et al. (2022). Skin $CD4^+$ Trm Cells Distinguish Acute Cutaneous Lupus Erythematosus from Localized Discoid Lupus Erythematosus/subacute Cutaneous Lupus Erythematosus and Other Skin Diseases. *J. Autoimmun.* 128, 102811. doi:10.1016/j.jaut.2022.102811

Zhu, C., Fei, W., Wang, W., Tang, L., Gao, J., and Zhou, F. (2021). Copy Number Variation Analysis of IL22 and LCE3C in Different Subtypes of Psoriasis in a Chinese Han Population. *Med. Sci. Monit.* 27, e934927. doi:10.12659/msm.934927

Zhu, H., Zhao, M., Chang, C., Chan, V., Lu, Q., and Wu, H. (2021). The Complex Role of AIM2 in Autoimmune Diseases and Cancers. *Immun. Inflamm. Dis.* 9, 649–665. doi:10.1002/iid3.443

Zuo, X., Sun, L., Yin, X., Gao, J., Sheng, Y., Xu, J., et al. (2015). Whole-exome SNP Array Identifies 15 New Susceptibility Loci for Psoriasis. *Nat. Commun.* 6, 6793. doi:10.1038/ncomms7793

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.

Copyright © 2022 Wang, Gao, Huang, Jeong, Ko, Shen, Chen, Zhong, Zou, Yu and Shen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 07 September 2022
DOI 10.3389/fgene.2022.1008582

Check for updates

OPEN ACCESS

EDITED BY
Yonghu Sun,
Shandong Provincial Hospital of Dermatology, China

REVIEWED BY
Zhilian Jia,
City of Hope National Medical Center, United States
Longli Kang,
Xizang Minzu University, China

*CORRESPONDENCE
Iouri Chepelev,
ichepelev@gmail.com
John B. Harley,
johnbharley@yahoo.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

RECEIVED 01 August 2022
ACCEPTED 18 August 2022
PUBLISHED 07 September 2022

CITATION
Saint Just Ribeiro M, Tripathi P, Namjou B, Harley JB and Chepelev I (2022), Haplotype-specific chromatin looping reveals genetic interactions of regulatory regions modulating gene expression in 8p23.1.
Front. Genet. 13:1008582.
doi: 10.3389/fgene.2022.1008582

COPYRIGHT
© 2022 Saint Just Ribeiro, Tripathi, Namjou, Harley and Chepelev. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Haplotype-specific chromatin looping reveals genetic interactions of regulatory regions modulating gene expression in 8p23.1

Mariana Saint Just Ribeiro[1†], Pulak Tripathi[1†], Bahram Namjou[1], John B. Harley[2,3*] and Iouri Chepelev[2,3*]

[1]Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States, [2]Research Service, US Department of Veterans Affairs Medical Center, Cincinnati, OH, United States, [3]Cincinnati Education and Research for Veterans Foundation, Cincinnati, OH, United States

A major goal of genetics research is to elucidate mechanisms explaining how genetic variation contributes to phenotypic variation. The genetic variants identified in genome-wide association studies (GWASs) generally explain only a small proportion of heritability of phenotypic traits, the so-called missing heritability problem. Recent evidence suggests that additional common variants beyond lead GWAS variants contribute to phenotypic variation; however, their mechanistic underpinnings generally remain unexplored. Herein, we undertake a study of haplotype-specific mechanisms of gene regulation at 8p23.1 in the human genome, a region associated with a number of complex diseases. The *FAM167A-BLK* locus in this region has been consistently found in the genome-wide association studies (GWASs) of systemic lupus erythematosus (SLE) in all major ancestries. Our haplotype-specific chromatin interaction (Hi-C) experiments, allele-specific enhancer activity measurements, genetic analyses, and epigenome editing experiments revealed that: 1) haplotype-specific long-range chromatin interactions are prevalent in 8p23.1; 2) *BLK* promoter and *cis*-regulatory elements cooperatively interact with haplotype-specificity; 3) genetic variants at distal regulatory elements are allele-specific modifiers of the promoter variants at *FAM167A-BLK*; 4) the *BLK* promoter interacts with and, as an enhancer-like promoter, regulates *FAM167A* expression and 5) local allele-specific enhancer activities are influenced by global haplotype structure due to chromatin looping. Although systemic lupus erythematosus causal variants at the *FAM167A-BLK* locus are thought to reside in the *BLK* promoter region, our results reveal that genetic variants at distal regulatory elements modulate promoter activity, changing *BLK* and *FAM167A* gene expression and disease risk. Our results suggest that global haplotype-specific 3-dimensional chromatin looping architecture has a strong influence on local allelic *BLK* and *FAM167A* gene expression, providing mechanistic details for how regional variants controlling the *BLK* promoter may influence disease risk.

KEYWORDS

gene regulation, enhancer, enhancer-like promoter, genetic interaction, missing heritability, *BLK*, systemic lupus erythematosus (SLE), *FAM167A*

Introduction

Contemporary genetics research aims to understand mechanisms of how variation in DNA sequence underlies phenotypic variation in normal and disease states. Genome-wide association studies (GWASs) have identified thousands of genetic loci associated with over 4,000 phenotypes (www.ebi.ac.uk/gwas/). However, genetic variants identified in GWAS studies explain only a modest fraction of total genetic risk in complex diseases, leading to the so-called missing heritability problem (Manolio et al., 2009). Recent evidence suggests that additional common variants beyond the lead GWAS single nucleotide polymorphisms (SNPs) contribute to disease risk (Gusev et al., 2013; Corradin et al., 2014; Gusev et al., 2014; Corradin et al., 2016; Boyle et al., 2017). However, mechanistic underpinnings of how additional genetic variants contribute to disease risk have not been thoroughly investigated for the vast majority of disease risk loci. Some suggest that a proper account of genetic interactions may help solve the missing heritability problem (Zuk et al., 2012). GWAS associations most often occur in non-coding regions of the genome, presumably at gene regulatory elements (GREs) (Hindorff et al., 2009), suggesting that additional risk variants will be found in GREs as well.

Transcriptional regulation is a complex process involving 3-dimensional (3D) chromatin interactions of GREs (Li et al., 2018). Combinations of genetic polymorphisms affecting various components in this process may thus alter gene expression and contribute to disease risk. The importance of 3D genome architecture for gene regulation in normal and disease states has been increasingly appreciated and is clearly a tremendous source for important new knowledge (Babu and Fullwood, 2015; Krijger and de Laat, 2016). The techniques based on chromosome conformation capture have emerged as methods of choice for mapping the 3D structure of the human genome (de Wit and de Laat, 2012). In particular, capture Hi-C, a high-throughput method to identify chromatin interactions in large genomic regions, has been used to determine high-resolution 3D genome structures at some disease-associated loci (Dryden et al., 2014; Jäger et al., 2015).

In order to investigate the role of haplotype-specific 3D chromatin structure in allelic gene expression, we have chosen the 8p23.1 region on human chromosome 8.8p23.1 contains associations with a number of complex diseases and the largest (4.5 Mb) known common DNA inversion region in humans (Salm et al., 2012; Namjou et al., 2014). The *FAM167A-BLK* locus in this region, for example, has been consistently found to be associated with systemic lupus erythematosus (SLE) GWASs in all major ancestries (International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN) et al., 2008; Hom et al., 2008; Ito et al., 2009; Suarez-Gestal et al., 2009; Sánchez et al., 2011; Castillejo-López et al., 2012; Delgado-Vega et al., 2012; Guthridge et al., 2014).

Associations with the *BLK* locus have also been identified in rheumatoid arthritis (RA), systemic sclerosis (SSc), Sjögren's syndrome (SjS), Kawasaki's disease (KD), antiphospholipid syndrome (APS) and maturity-onset diabetes of the young (MODY), thus justifying a focused mechanistic study of this locus (Hom et al., 2008; Borowiec et al., 2009; Yin et al., 2009; Gourh et al., 2010; Tsuchiya et al., 2010; Lessard et al., 2013). BLK is a signal transduction molecule that is important for sustaining the inflammatory response, including autoimmune responses. BLK probably plays an important role in early B cell development; thus, its dysregulation may result in a breakdown of peripheral self-tolerance during B cell development (Nashi et al., 2010; Simpfendorfer et al., 2012).

SLE patients have lower levels of the *BLK* gene product, a finding that is correlated with the risk alleles at the *FAM167A-BLK* locus (Hom et al., 2008). The strongest SLE association signal at the *BLK* locus is from the promoter SNP rs13277113 (Hom et al., 2008). The trans-population mapping and sequencing strategy was used in (Guthridge et al., 2014) to identify two putative SLE causal variants, rs922483 and rs1382568, at the *BLK* promoter region. The data available have led to a general consensus in the SLE genetics research community that causal variants at *FAM167A-BLK* are very likely to be in the *BLK* promoter region.

Our haplotype-specific chromatin interaction high-resolution Hi-C experiments, allele-specific enhancer activity H3K27ac measurements, genetic analyses, and epigenome editing experiments revealed that 1) haplotype-specific long-range chromatin interactions in 8p23.1 are prevalent, 2) *BLK* promoter and *cis*-regulatory elements cooperatively and haplotype-specifically interact, 3) *BLK* promoter interacts with and, as an enhancer-like promoter, regulates *FAM167A* expression, and 4) genetic variants at distal regulatory elements are allele-specific genetic modifiers of the promoter variants at *FAM167A-BLK*.

Our findings suggest a 'risk dosage' model whereby disease risk alleles at multiple regulatory elements at *BLK* locus synergistically decrease gene expression, thereby increasing SLE disease risk. Although SLE causal genetic variants are thought to reside in the *BLK* promoter region, we show that haplotype-specific distal genetic variants at regulatory elements modulate the effects of *BLK* promoter variants on gene expression and disease risk. More generally, our results suggest that global haplotype-specific 3-dimensional chromatin looping architecture may have a strong influence on local allelic gene expression and disease risk in SLE, as

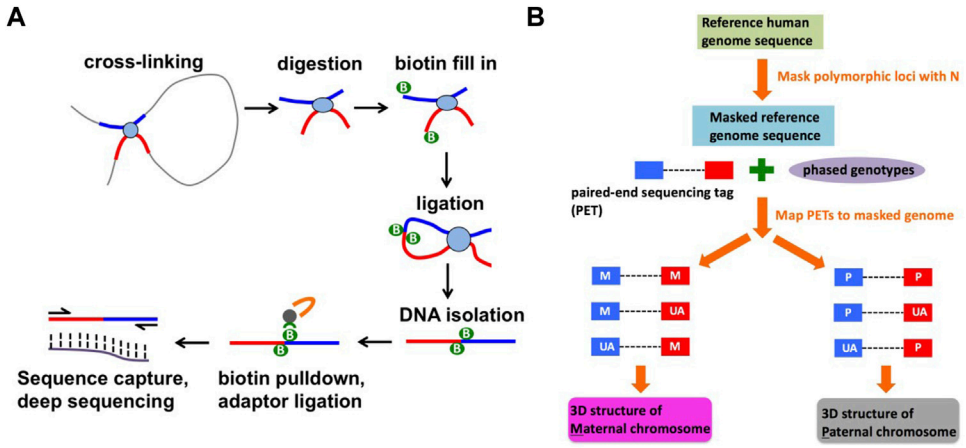


FIGURE 1
Identification of haplotype-specific chromatin interactions. **(A)** A schematic diagram of capture Hi-C method. **(B)** A computational strategy to identify haplotype-specific chromatin interactions.

well as in other complex diseases, and provide a model for a risk dosage approach to regulatory susceptibility loci, in general.

Results

Haplotype-specific chromatin looping

Long-range chromatin interactions are important for transcriptional regulation of genes (Chepelev et al., 2012; Krijger and de Laat, 2016). We suspected that causal genetic variants at *FAM167A-BLK* dysregulate normal gene expression by altering chromatin looping interactions. We thus hypothesized that in B cell lines heterozygous for SLE-associated SNPs in the 8p23.1 region, 3D chromatin structures of risk and non-risk haplotypes would differ. To test this hypothesis, we prepared capture Hi-C libraries from two 1,000 Genomes EBV-infected lymphoblastoid cell lines (LCLs), NA07000 and NA07056, both heterozygous for the SLE-associated SNP rs922483 (Guthridge et al., 2014), located in the promoter of *BLK*.

Capture Hi-C is a high-throughput, cost-effective method to identify long-range 3D chromatin interactions in a subset of the genome (Dryden et al., 2014) (see Materials and Methods, and Figure 1). We identified the haplotype-resolved 3D structure of a 3 Mb region in 8p23.1 at 5 kb resolution. At the false discovery rate (FDR) of 5%, we have identified 780 and 791 differential haplotype-specific chromatin interactions in NA07000 and NA07056 cell lines, respectively (Supplementary Tables S1, S2). Approximately half of these interactions are stronger and half of the interactions are weaker on SLE risk haplotype than on non-risk haplotype. Among the differential haplotype-specific interactions in 8p23.1, some are in the vicinity of the *BLK* locus (Figure 2).

In the ENCODE Project (encodeproject.org) data, several extended regions at the *BLK* locus show enrichment for the H3K27ac histone modification mark in LCLs and primary B cells. (We designate these regions E1, E2, E3, and P in Figure 2). The H3K27ac mark is thought to be associated with active enhancers in the genome (Creyghton et al., 2010). Our haplotype-specific Hi-C data described below strongly suggest that these regions positively regulate *BLK* expression. Our capture Hi-C data show that enhancer E3, located more than 50 kb away from the *BLK* transcription start site, interacts with the promoter in a haplotype-specific manner. Consistent with lower expression of *BLK* on the SLE risk haplotype, P-E3 interaction frequency on the SLE risk haplotype is lower than on the non-risk haplotype (Figure 2). It has been known for some time that enhancers can interact with other enhancers in the 3D genome (Chepelev et al., 2012). We found that enhancers also interact with each other in a haplotype-specific way (Figure 2), again with weaker interactions on the risk haplotype. These results are consistent with long-range chromatin interactions between enhancers playing a role in co-operatively enhancing *BLK* expression.

Intriguingly, we found that the interaction of the *BLK* promoter with a distal REST binding site located 115 kb away (Figure 2) is stronger on SLE risk haplotype, consistent with the repressive role of REST (repressor element 1-silencing transcription factor) (Ooi and Wood, 2007) and lower expression level of *BLK* gene on SLE risk haplotype.

Function of the *FAM167A* gene located 27 kb upstream of *BLK* has recently been shown to be an activator of the non-canonical activation pathway of NFκB, which is potentially important for mechanisms of the inflammatory phenotypes

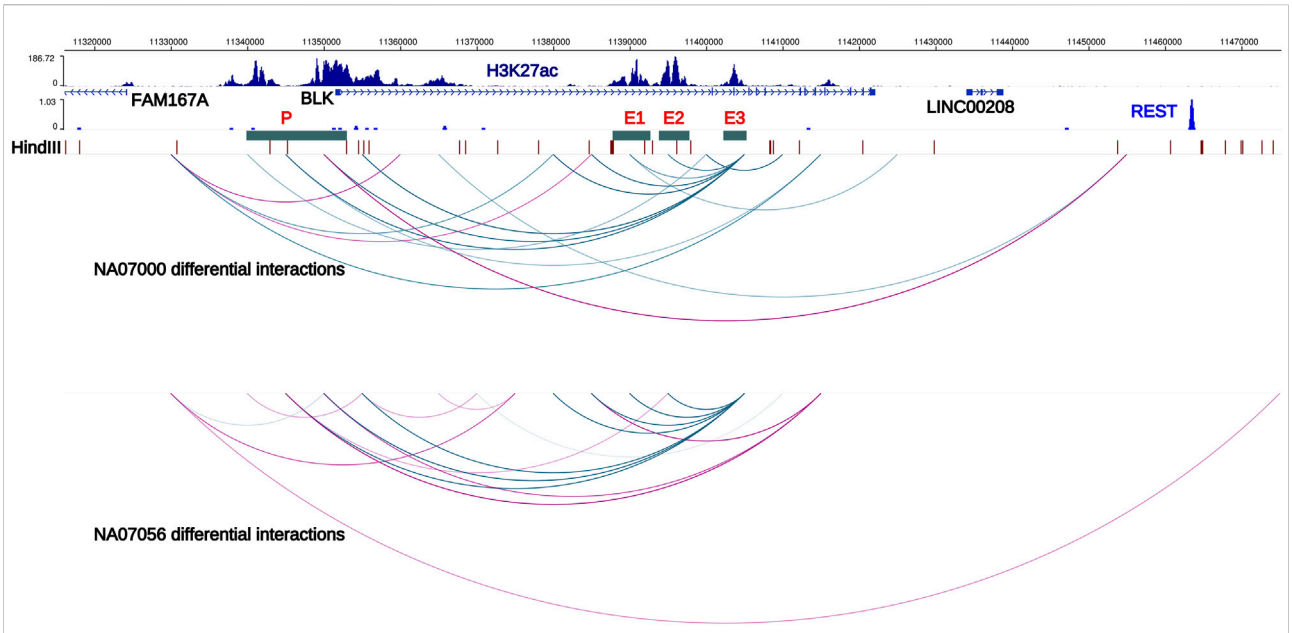


FIGURE 2
Haplotype-specific chromatin interactions at *FAM167A-BLK* locus. Arcs represent haplotype-specific chromatin interactions from our Hi-C data in NA07000 and NA07056 LCLs. Blue arcs depict long-range chromatin interactions which are weaker on risk haplotype. Purple arcs depict interactions that are stronger on the risk haplotype than on the non-risk haplotype. Interaction of *BLK* promoter with the enhancer E3 is weaker and with the REST repressor binding region is stronger on SLE risk haplotype, consistent with reduced expression of *BLK* on the risk haplotype. Interaction of enhancer E3 with enhancers E1 and E2 is weaker on the risk haplotype. Also displayed are the H3K27ac signals in NA12878 LCLs (ENCODE data) and HindIII restriction sites. E1, E2, E3; putative enhancers 1, 2, and 3. P; *BLK* promoter, locus of SLE credible set BLK-CS-11 (Supplementary Table S5).

with genetic associations at this locus (Mentlein et al., 2018; Yang et al., 2022). We, therefore, retain *FAM167A* as an SLE candidate gene, acknowledging the existing circumstantial evidence supporting *BLK* as a participant in the mechanism altering disease risk. *FAM167A* is upregulated 8-fold upon B cell receptor stimulation (publicly available RNA-seq data in GEO GSE61608). Further, RA-associated variants exhibit high LD (r^2 > 0.8) with a B cell selective cis-eQTL for *FAM167A* expression identified in a cohort of early RA patients (Thalayasingam et al., 2018). The haplotype-specific expression patterns of these two genes are anti-correlated: *BLK* is less expressed on the SLE risk haplotype than on the non-risk haplotype, whereas *FAM167A* is more expressed on the risk haplotype than on the non-risk haplotype (Figure 3B) (Hom et al., 2008). Our capture Hi-C data revealed that *FAM167A* and *BLK* promoters interact, with the interaction being stronger on risk-haplotype in NA07000, but not in NA07056 cells (Figures 2,3A). Our model is summarized in Figure 3B. It has previously been proposed that promoters can regulate other promoters *via* chromatin looping interaction (Xu et al., 2011). Many promoter-promoter chromatin interactions have previously been identified genome-wide (Chepelev et al., 2012; Li et al., 2012). Promoters with enhancer activity, the so-called enhancer-like promoters, have been described recently (Dao et al., 2017; Dao and Spicuglia, 2018). That the promoters of *FAM167A* and *BLK* genes may functionally interact in this manner is consistent with these data and remains an intriguing possibility.

Since SLE causal variants at *FAM167A-BLK* are very likely to be in the *BLK* promoter region (see discussion in Introduction) and given our findings above, we hypothesized that these variants may have functional effects on modulating *FAM167A* expression, and on disease risk, via haplotype-specific chromatin interaction between promoters of *BLK* and *FAM167A*. We thus sought to perturb *BLK* expression and measure changes in *FAM167A* expression.

We infected NA07000 LCL cells, in which *BLK* and *FAM167A* are expressed, with a lentivirus to express a dCas9-KRAB fusion protein for targeted gene repression (see Figure 4A) (Thakore et al., 2015). When localized to genomic DNA, KRAB recruits a heterochromatin-forming complex that causes histone methylation and deacetylation. We targeted dCas9-KRAB to the E3 enhancer region in NA07000 cells by transfection of plasmids expressing appropriate guide RNAs (gRNAs) (Supplementary Tables S3,S4). The *BLK* expression was down-regulated by ~19% and *FAM167A* expression was up-regulated by ~37% as a consequence of the targeting of dCas9-KRAB to the E3 enhancer region (Figures 4B,C).

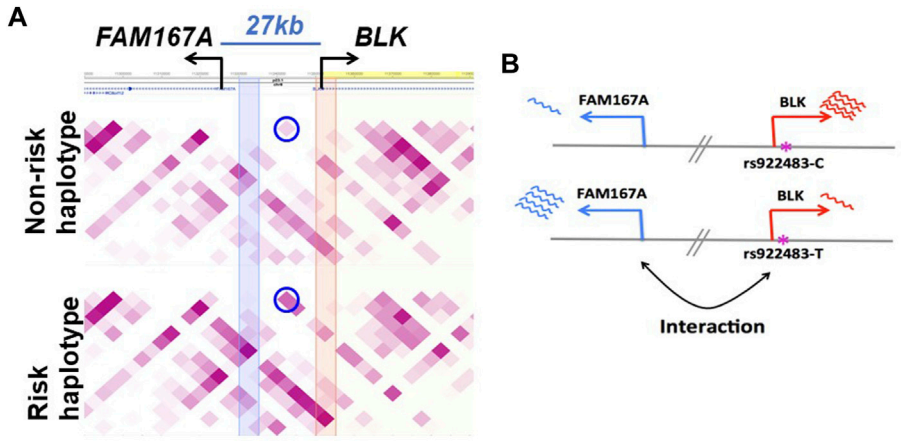


FIGURE 3
Haplotype-specific *FAM167A-BLK* promoter-promoter interactions in NA07000 cells. **(A)** *FAM167A-BLK* interaction is stronger on SLE risk haplotype. Shown is a heatmap view of Hi-C chromatin interaction frequencies on non-risk and risk haplotypes, darker purple representing higher interaction frequencies. Each square represents an interaction between two 5 kb genomic regions located at the diagonal extension intersecting the horizontal line. The circled squares represent interactions between two 5 kb regions located where blue and orange vertical bands meet the horizontal line (at *FAM167A* and *BLK* promoters). **(B)** *BLK* and *FAM167A* have anti-correlated expression patterns. *BLK* is less and *FAM167A* is more expressed on the risk (lower panel) than on the non-risk (upper panel) haplotype. Shown also is a putatively SLE causal promoter SNP rs922483 with its non-risk (C) and risk (T) alleles.

The up-regulation of *FAM167A* may be a direct or an indirect consequence of perturbation of chromatin state at enhancer E3 by KRAB. In order to distinguish between these two possibilities, we chose K562 cells to perform epigenome editing to activate enhancer E3. The *BLK* and *FAM167A* genes are not expressed beyond the basal level in K562 cells and the *BLK* enhancers are not active since they have no H3K27ac signal in K562 cells (Supplementary Figure S1). We infected K562 cells with a lentivirus to express a dCas9-p300 fusion protein (Hilton et al., 2015; Klann et al., 2017) (see Figure 6A). The p300 is a histone acetyltransferase that acetylates H3K27. While *BLK* is strongly up-regulated by targeting of dCas9-p300 to enhancer E3 in K562 (Figure 4B), *FAM167A* expression did not change (Figure 4C). This suggests that the E3 region is not directly involved in regulating *FAM167A* expression. More likely, the effect of targeting dCas9-KRAB to E3 on *FAM167A* expression is indirect and is mediated by increased long-range chromatin interaction frequency between promoters of *BLK* and *FAM167A*, brought about by down-regulation of *BLK* expression (see the model in Figure 4D). We hypothesize that upon down-regulation of *BLK* transcription, an enhancer-like activity of *BLK* promoter goes up, which leads to its long-range chromatin interaction with *FAM167A* promoter and up-regulation of *FAM167A* expression.

Haplotype-specific enhancer activity

Our Hi-C data revealed haplotype-specific interactions of enhancers with the *BLK* promoter. We thus hypothesized that enhancer activities should be stronger on the SLE non-risk haplotype. We cloned NA07056 risk and non-risk haplotype enhancer sequences into luciferase reporter vectors and transfected LCLs with these constructs. The reporter assay data show that the activity of enhancer E3 is haplotype-specific, and is less active on the SLE risk haplotype (Figure 5A). This result is consistent with the chromatin interaction analysis above. Reporter experiments revealed that enhancer activity of non-risk E3 sequence is ~1.5 times higher than the activity of risk E3 sequence (Figure 5A). Intriguingly, our allele-specific ChIP-qPCR experiments showed that the H3K27ac signal at E3 enhancer on non-risk haplotype is ~8.5 times higher than on the risk haplotype (Figure 5B). We hypothesize that the allelic differences in regulatory element activities are determined not only by local DNA sequence variations at the element but also by allelic differences at distal sites, due to the amplification effect of chromatin looping interactions in the 3D chromatin context. We have tested our hypothesis for one enhancer (E3) in one cell line in our experiments, suspecting that this is a more general phenomenon and that analogous findings would hold true also at many other regulatory regions.

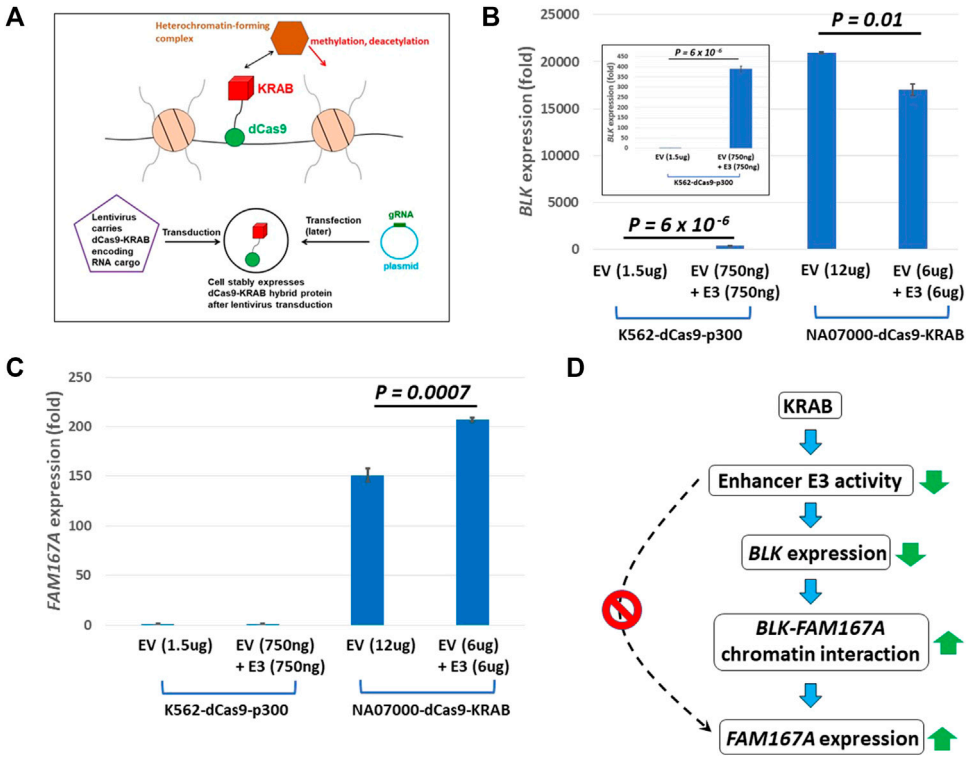


FIGURE 4
dCas9-KRAB targeting reveals regulation of *FAM167A* expression by *BLK* promoter region. **(A)** Schematics of epigenome editing (repressor KRAB) experiment in NA07000 LCL cell line. **(B)** Targeting of dCas9-KRAB to enhancer E3 in NA07000 cells results in down-regulation of *BLK*. Targeting of dCas9-p300 to enhancer E3 in K562 cells results in ~400 folds up-regulation of *BLK* (two left-most bars in the main figure and the inset for an expanded view). For the schematics of dCas9-p300 experiments, see Figure 6A. **(C)** Targeting of dCas9-KRAB to enhancer E3 in NA07000 cells results in up-regulation of *FAM167A* expression. **(D)** A model to explain data in **(B–C)**. The dotted arrow with the red 'STOP' sign symbolizes the following finding from **(C)** (two left-most small bars grouped as 'K562-dCas9-p300' cell line): targeting of activator p300 to enhancer E3 in K562 cells did not result in up-regulation of *FAM167A* expression. This suggests that E3 does not directly regulate *FAM167A* expression. Rather, the effect of repressing E3 on *FAM167A* expression is mediated by *FAM167A-BLK* promoter-promoter interaction (see Figures 2,3). In **(B,C)**, the *BLK* and *FAM167A* expression values shown are relative to their expression levels in K562-dCas9-p300 cells transfected with empty vector (EV) (the left-most bars in **(B)** and **(C)** have values identically equal to 1). EV, empty vector.

Epigenome editing reveals participating elements in *BLK* regulation

Our Hi-C data revealed haplotype-specific chromatin looping interactions between enhancers and the promoter at the *BLK* locus. Based on these findings, we hypothesize that enhancers and promoter pairs synergistically regulate *BLK* expression. For the lack of baseline *BLK* expression and the absence of an H3K27ac signal at the enhancers (Supplementary Figure S1), we again chose K562 cells, here performing epigenome editing to activate *BLK* expression and test the hypothesis. We infected K562 cells with a lentivirus to express a dCas9-p300 fusion protein (Hilton et al., 2015; Klann et al., 2017) (see Figure 6A). The p300 is a histone acetyltransferase that acetylates H3K27. We targeted dCas9-p300 to E2, E3 and promoter regions singly and in combinations in K562 cells by transfection of plasmids expressing appropriate guide RNAs (gRNAs) (Supplementary Tables S3,S4). Strong synergistic activation of the *BLK* gene (500 to 900-fold increases) was observed for E2-E3 and E3-promoter targeting by p300 (Figures 6B,C), thereby confirming the regulatory role of this genome region for *BLK*.

Multiple genetic variants contribute to variance in *BLK* expression

Since gene regulation is a complex process involving interactions of multiple regulatory regions and protein complexes, we reasoned that many genetic variants may contribute to variance in *BLK* expression. We analyzed gene expression and genotype data from 344 European individuals (Lappalainen et al., 2013), henceforth denoted as dataset D344, using the genetic relationship matrix approach as implemented

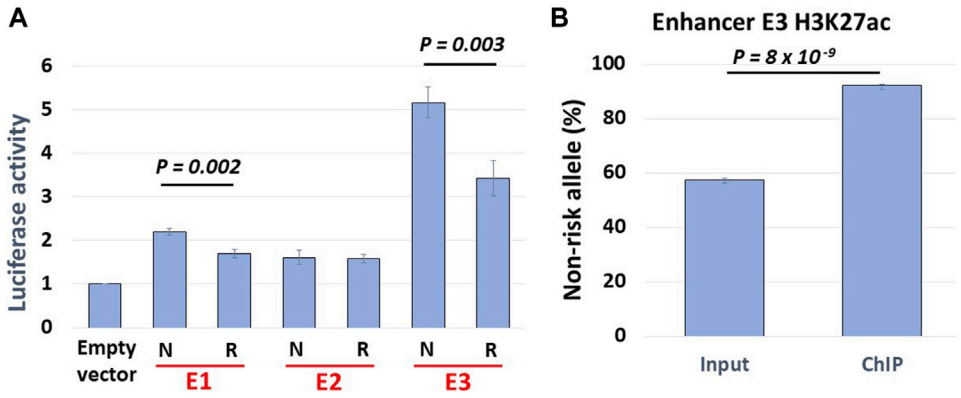


FIGURE 5
Haplotype-specific enhancer activity. **(A)** Allelic luciferase activities of enhancers in NA07056 cells. N: SLE non-risk haplotype sequence, R: SLE risk haplotype sequence. Non-risk enhancer E3 sequence reporter activity is ~1.5x higher than that of risk sequence. **(B)** Allele-specific ChIP-qPCR at E3 enhancer in NA07056 cells in which the SNP rs2244931 is heterozygous. Non-risk to risk enhancer activity ratio in native chromatin context is significantly higher ~8.5x (calculated as normalized ratio $(N/R)_{ChIP}/(N/R)_{Input}$; See Materials and Methods for details), presumably due to "amplification of E3 enhancer activity due to long-range chromatin interactions" effect (see haplotype-specific E1-E3 and E2-E3 interactions in Figure 2).

in GCTA software (Yang et al., 2010) to estimate *BLK* gene expression variance explained by various sets of SNPs. The SLE-associated proxy rs922483 alone explained around 25% of expression variance. Inclusion of all SNPs in putative regulatory regions of *BLK*, as defined by the H3K27ac epigenetic mark and REST binding region (see Figure 2; Supplementary Table S7), increased the variance explained to 44%, supporting the contention that multiple SNPs at the *BLK* locus explain variance in gene expression.

Enhancer haplotypes modulate *BLK* expression

Given our findings, we sought to find more direct evidence for the causal effects of enhancer haplotypes on *BLK* expression. To this end, we derived an SLE credible set of 11 SNPs, termed BLK-CS-11, from existing genotype data (see Materials and Methods, and Supplementary Table S5) and used phased genotype data for individuals of European ancestry from the 1000 Genomes Project (1000genomes.org) to extract haplotype sequences of the *BLK* locus. We then computed frequencies of different promoter haplotype sequences in the dataset.

The frequency of promoter risk haplotype CCCCTTAAACA (based on 11 SNPs in BLK-CS-11) (Figure 7A), which we denote henceforth as prom-R, is 24%. The most frequent non-risk haplotype TGTACCGGGTG, which we denote as prom-N, is present at 72%. In subsequent analyses, we will focus on these two main haplotypes, which together account for 96% of all European haplotypes. The rows in the heatmap plot in Figure 7A correspond to the haplotypes, with the rows labeled as "Risk" representing promoter "risk" haplotypes and the rows labeled as 'Non-risk' representing promoter "non-risk" haplotypes, respectively.

We next asked, conditioned on the promoter haplotype, what are the local haplotype sequences at candidate enhancer E1, E2, and E3 regions defined by islands of H3K27ac enrichment? The promoter and enhancer haplotype sequences are displayed in Figure 7A. The haplotypes are grouped based on promoter haplotype: risk for prom-R (upper block) and non-risk for prom-N (lower block). . We conditioned on the promoter haplotype prom-R and counted different local haplotypes at E1, E2, and E3. The most frequent E1, E2, and E3 haplotypes, conditioned on prom-R haplotype, denoted as E1-R, E2-R, and E3-R, respectively, are present in 87%, 85%, and 54% of all prom-R haplotypes, respectively. Our choice of the label "R" in the naming of these enhancer haplotypesmay seem arbitrary; the use of this label emphasizes the effect of these haplotypes on *BLK* expression and disease risk discussed below. Interestingly, the haplotype combination (prom-R, E1-R, E2-R, E3-R), which can be seen in the upper dotted box in Figure 7A, is frequent and constitutes 48% of all promoter-E1-E2-E3 haplotypes conditioned on prom-R and 11% of all European subject chromosomes.

Similarly, the most frequent E1, E2, and E3 haplotypes, conditioned on prom-N haplotype, denoted as E1-N, E2-N, and E3-N, respectively, are present in 25%, 70%, and 50% of all prom-N haplotypes, respectively. Interestingly, the haplotype combination (prom-N, E1-N, E2-N, E3-N), which can be seen in the lower dotted box in Figure 7A, is frequent and constitutes 24% of all promoter-E1-E2-E3 haplotypes conditioned on prom-N. All other alternative local haplotypes at E1, E2, and E3,

different from N and R haplotypes defined above, are denoted by E1-A, E2-A, and E3-A, respectively. We will use simplified notation, such as RNRN for (prom-R, E1-N, E2-R, E3-N) haplotype, subsequently.

We next asked: how does the *BLK* mRNA level depend on different haplotype combinations at the promoter, E1, E2, and E3? To answer, we used RNA-seq expression data from dataset D344 [gene expression from 344 European individuals (Lappalainen et al., 2013)]. We have defined two groups of samples based on their multilocus genotypes (i.e., pairs of haplotypes in each individual) NNNN/NNNN and RRRR/RRRR. We have also defined two more groups of samples based only on promoter multilocus genotypes prom-N/prom-N and prom-R/prom-R, denoted N/N and R/R, respectively. The boxplot of expression levels in these 4 groups is shown in Figure 7B. There is a clear "risk-dosage" dependent decrease in *BLK* expression level in these data, with the contributions of "risk" from both the promoter and enhancer variants. Analyses for *BLK* promoter-enhancer pairs, shown in Figures 7C–E, support a risk-dosage model according to which the risk status of local haplotypes at cis-regulatory regions cumulatively contributes to a reduction in *BLK* expression. As defined earlier, alternative local haplotypes at E1, E2, and E3, different from N and R haplotypes, are denoted as "A". By replacing N's with A's in the analyses, we found almost identical results as above, suggesting that A and N enhancer haplotypes have similar effects on *BLK* expression (data not shown).

Genetic interactions at *FAM167A-BLK* locus are enriched in regulatory regions

We have provided evidence supporting the hypothesis that enhancer haplotypes can influence *BLK* expression by acting as genetic modifiers of promoter haplotypes. However, we excluded from our analysis genetic variants outside of gene regulatory elements/regions (GREs). What is the contribution of the non-GRE variants to *BLK* expression and do these variants interact? To address this question, we analyzed data D344 [gene expression data from 344 European individuals (Lappalainen et al., 2013)]. For each pair of SNPs, we fitted a linear ($y = b_0 + b_1x_1 + b_2x_2$) and an interaction ($y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2$) model to *BLK* expression and genotype data (y denotes expression level and x_1, x_2 denote 0/1/2 coded genotypes of the two SNPs under consideration). The log-likelihood ratio test was used to determine statistically significant genetic interactions. If our hypothesis is true, then the genetic interactions are predicted to be predominantly between SNPs located at GREs. However, SNPs at the *BLK* locus are in high linkage disequilibrium with each other; therefore, the statistically identified genetic interactions may not necessarily be between genetic variants that causally influence *BLK* expression.

In order to perform an unbiased test of genetic interaction enrichment at regulatory regions, we separated genetically interacting SNP pairs identified by our statistical analysis into two groups: 1) SNPs pairs that are in two different gene regulatory regions (GRE-GRE group) and 2) SNP pairs with at least one SNP not in a regulatory region (non-GRE-GRE group). The genomic coordinates of gene regulatory regions and details of the analysis are given in Materials and Methods. A ranked list of $-\log_{10}(p\text{-value})$ of genetic interactions in each group was generated. Quantile-quantile analysis revealed that genetic interactions in the GRE-GRE group are more significant than in the non-GRE-GRE group (p-value = 0.0003) (Figure 8), thereby demonstrating enrichment of genetic interactions between gene regulatory elements.

Materials and methods

Cell culture

K562 cells were obtained from the American Tissue Collection Center (ATCC). NA07056 and NA07000 LCLs were obtained from the Coriell Institute for Medical Research cell repository. K562 cells were maintained in Iscove's Modified Dulbecco's Medium supplemented with 10% FBS and 1% penicillin-streptomycin. LCLs were maintained in RPMI medium supplemented with 10% FBS and 1% penicillin-streptomycin. K562 and LCLs were grown at 37°C in 5% CO_2.

Capture Hi-C experiments

Capture Hi-C libraries were prepared according to the protocol described in (Jäger et al., 2015). The protocol consists of two parts: Hi-C library preparation and target enrichment (Figure 1A). A SureSelect Custom Target Enrichment Library covering a 3 Mb region in the 8p23.1 (hg19 coordinates: chr8:8,190,000–11,838,000) was designed using eArray software (Agilent). Hi-C library preparation, comprising chromatin fixation, HindIII digestion, biotin labelling, ligation, and crosslink reversal was performed as described in (Rao et al., 2014) with minor modifications described in (Jäger et al., 2015). Target enrichment was performed according to the SureSelect protocol (Agilent) with minor modifications described in (Jäger et al., 2015). We have prepared 10 capture Hi-C libraries from 5 independent batches of NA07000 cells and 4 capture Hi-C libraries from 2 independent batches of NA07056 cells. The libraries were sequenced on Illumina HiSeq 2,500 system, producing 461 million and 206 million

paired-end 2 × 125 bp reads for NA07000 and NA07056, respectively.

Allele-specific luciferase assay

We amplified ~2 kb DNA fragments located in the E1, E2, and E3 enhancer regions (see Figure 2; the hg19 coordinates of the fragments are chr8:1,1,391,971–11,391,971, chr8: 11,394,297–11,396,127 and chr8:11,402,768–11,404,795, respectively) from the genomic DNA of NA07056 lymphoblastoid cells using PCR with primers listed in Supplementary Table S6. The PCR products were cloned into pCR2.1-TOPO vector (catalog #K4500-01, Invitrogen) and sub-cloned into pGL4 luciferase reporter vectors (catalog #E6651, Promega). The bacterial cells were then transformed, and single-cell colonies were isolated and Sanger-sequenced to identify SLE risk and non-risk clones using phased haplotype data from the 1000 Genomes Project.

An internal control reporter vector containing Renilla luciferase was simultaneously transfected with our experimental vectors as a control for assay-to-assay variability. One microgram of each vector was transfected into the NA07056 (10^6 cells per sample in triplicate). Cells were then incubated at 37°C for 24 h. Luciferase activity was measured with the Dual-Luciferase Reporter Assay System (catalog #E1960, Promega). Luciferase activity was normalized through the division of *BLK* risk or non-risk construct reporter activity by the reporter activity of the pRL-TK-Renilla luciferase construct. The mean and standard error of measurement were calculated on the basis of the normalized luciferase activities. The one-sided Student's t-test was used to compare the N and R groups in Figure 5A.

Allele-specific H3K27ac ChIP-qPCR

Three independent chromatin immunoprecipitation (ChIP) experiments in NA07056 cells were carried out using antibodies against H3K27ac (catalog #C15410196, pAb-196–050, Diagenode) following a standard protocol. Enhancer E3 (see Figure 2) SNP rs2244931 is heterozygous in NA07056 (C/G, where G is on the SLE risk haplotype). Allelic H3K27ac levels at rs2244931 were quantified using TaqMan custom SNP Genotyping Assay, TaqMan Genotyping Master Mix (catalog # 4,371,353, Applied Biosystems), and input/ChIP DNA. The allelic ratios in the ChIP and input DNA were determined by fitting log2 transformed VIC/FAM ratios to a standard curve constructed from DNA with known rs2244931 allelic ratios obtained by mixing gRNA from cell lines homozygous for rs2244931 (C/C and G/G) as well as heterozygous (C/G). The normalized non-risk to risk allele ratio of the H3K27ac signal was determined as the ratio $(N/R)_{ChIP}/(N/R)_{Input}$ of non-risk/risk allele ratios in ChIP and input DNA samples. In Figure 5B, the percentage of the non-risk allele RT-qPCR signal in input DNA is ~57.5% and that of the risk allele is ~42.5%. Similarly, the non-risk allele RT-qPCR signal in ChIP DNA constitutes ~92% and the risk allele signal constitutes ~8%. Thus, the normalized (true) non-risk to risk ratio of the H3K27ac signal is (92/8)/(57.5/42.5) = 8.5. The one-sided Student's t-test was used to compare percentages of non-risk alleles RT-qPCR signals in input and ChIP DNA (Figure 5B).

CRISPR epigenome editing

The plasmid pLV-dCas9-p300-P2A-PuroR was a gift from Charles Gersbach (Addgene plasmid # 83,889) (Klann et al., 2017). The plasmid pLV hU6-sgRNA hUbC-dCas9-KRAB-T2a-GFP was a gift from Charles Gersbach (Addgene plasmid # 71,237) (Thakore et al., 2015). The plasmid pSPgRNA was a gift from Charles Gersbach (Addgene plasmid #47108) (Perez-Pinera et al., 2013). For dCas9-p300 and dCas9-KRAB experiments, lentivirus was produced by transfecting HEK293 cells with pLV-dCas9-p300-P2A-PuroR and pLV hU6-sgRNA hUbC-dCas9-KRAB-T2a-GFP plasmids, respectively, at Cincinnati Children's Hospital Medical Center's Viral Vector Core facility.

For dCas9-p300 experiments, 5 million K562 cells were incubated for 3 days with the concentrated dCas9-p300 lentivirus at the cell to virus ratio of 1:5 in the presence of 8ug/ml polybrene. After 3 days, virus-infected cells were selected with 2 ug/ml of puromycin. We named the resulting cells K562-dCas9-p300.

For dCas9-KRAB experiments, 10 million NA07000 cells were spinfected (in 12 well plate, 1 million cells/ml of media, centrifuged at 800Xg for 2 h at 30°C) with the concentrated dCas9-KRAB lentivirus at the cell to virus ratio of 1:5 in the presence of 8 ug/ml polybrene. After spinfection, 1 ml of fresh media was added to each well (2 ml total) and cultured for 4 days at 37°C. Cells were then washed 3 times with fresh cultured media and continued to grow for 10 more days. After 2 weeks of infection, GFP + virus-infected cells were sorted and cultured to grow more GFP + virus-infected cells. We named the resulting cells NA07000-dCas9-KRAB.

For the regions P, E2 and E3 in Figure 2, we designed several gRNAs targeting these regions (see Supplementary Table S3 for the list of gRNA oligos and Supplementary Table S4 for Addgene IDs of the plasmids generated in this study). Each gRNA oligo pair was phosphorylated using T4 PNK, annealed and cloned into BbsI-digested pSPgRNA plasmid. For each of the P, E2, and E3 regions, we generated equimolar pools of gRNA plasmids (see Supplementary Tables S3,S4).

For *BLK* repression experiments, NA07000-dCas9-KRAB cells were transfected with E3 gRNA plasmid together with the empty vector (EV + E3) or the empty vector alone (EV) (see Figure 4). The empty vector is simply the intact pSPgRNA plasmid. As indicated in Figure 4, the total amount of plasmids

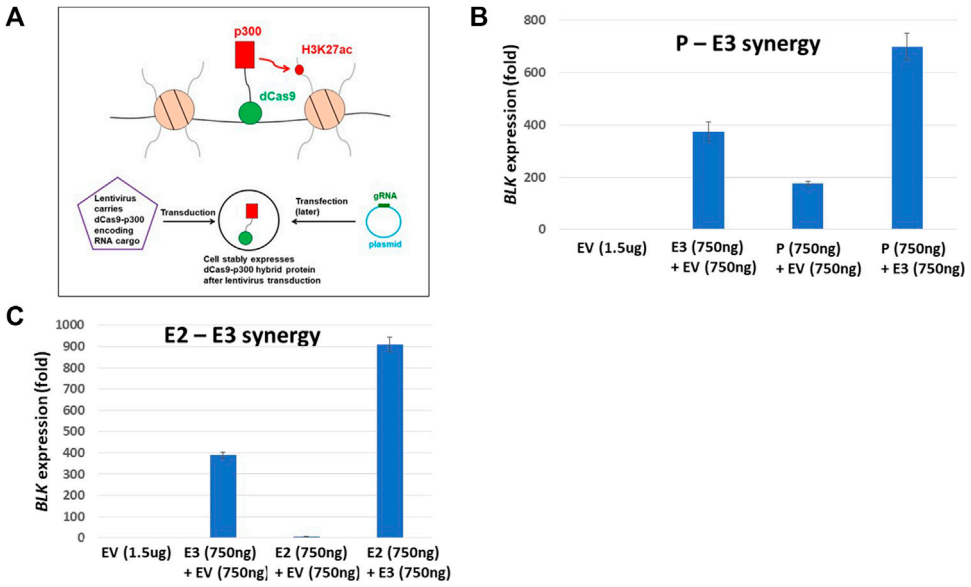


FIGURE 6
dCas9-p300 targeting reveals synergistic interaction of regulatory elements. **(A)** Schematics of epigenome editing (activator p300) experiment in K562 cell line. **(B)** P-E3 synergy: upregulation of *BLK* expression when both the *BLK* promoter P and the enhancer E3 are targeted by p300 protein is significantly higher than when p300 is targeted to either P (p-value = 0.004) or E3 (p-value = 0.004) alone. **(C)** E2-E3 synergy: upregulation of *BLK* expression when both enhancers E2 and E3 are targeted by p300 protein is significantly higher than when p300 is targeted to either E2 (p-value = 0.0008) or E3 (p-value = 0.0008) alone. In B-C, the *BLK* expression values shown are relative to the *BLK* expression level in K562-dCas9-p300 cells transfected with empty vector (EV) (the left-most bars in **(B)** and **(C)** have values identically equal to 1).

used in each transfection was 12 ug, with 6 ug of EV plasmid included whenever it was necessary in order to have a balanced total DNA amount of 12 ug. The optimal amount, 12 ug, of plasmids used for transfection in NA07000-dCas9-KRAB experiments above was determined from exploratory transfection experiments performed with E3 gRNA plasmid alone at varying DNA amounts and selecting the amount which resulted in the largest down-regulation of *BLK* expression (data not shown).

For synergy experiments, K562-dCas9-p300 cells were transfected with pairs of gRNA pools (P + E3 or E2+E3), a single gRNA pool together with the empty vector (EV + P, EV + E2, or EV + E3) or the empty vector alone (EV) (see Figure 6). The empty vector is simply the intact pSPgRNA plasmid. As indicated in Figure 6, the total amount of plasmids used in each transfection was 1.5 ug, with 750 ng of EV plasmid included whenever it was necessary in order to have a balanced total DNA amount of 1.5 ug. The optimal amount, 1.5 ug, of plasmids used for transfection in K562-dCas9-p300 experiments above was determined from exploratory transfection experiments performed with E3 gRNA plasmid alone at varying DNA amounts and selecting the amount which resulted in the highest upregulation of *BLK* expression (data not shown).

At 24 h, cells were harvested and mRNA was extracted using Dynabeads mRNA DIRECT kit (catalog # 61,006, Invitrogen). cDNA was generated using SuperScript™ IV VILO™ Master Mix kit (catalog # 11,756,050, Invitrogen). The *BLK* and *FAM167A* mRNA levels were measured by RT-qPCR using TaqMan Fast Advanced Master mix (catalog# 4,444,963, ThermoFisher Scientific) and TaqMan probes for *BLK* (Hs01017458_m1), *FAM167A* (Hs00697562_m1) and *GAPDH* (Hs03929097_g1). The one-sided Student's t-test was used to compare expression levels in Figures 4B,C and Figures 6B,C.

Analysis of capture Hi-C data

The capture Hi-C paired-end sequencing reads were mapped to the SNP-masked hg19 human genome using HiC-Pro software as follows (Servant et al., 2015) (Figure 1B). First, we obtained phased genotype VCF files for NA07000 and NA07056 cell lines from the 1000 Genomes Project (version v5a.20,130,502), and generated two SNP-masked hg19 genomes files for these cell lines. The heterozygous (Ref/Alt) and homozygous alternative allele (Alt/Alt) loci in the genomes were masked as "N" in order to mitigate the reference mapping bias. The individual reads in each mapped paired-end read are evaluated for the presence of

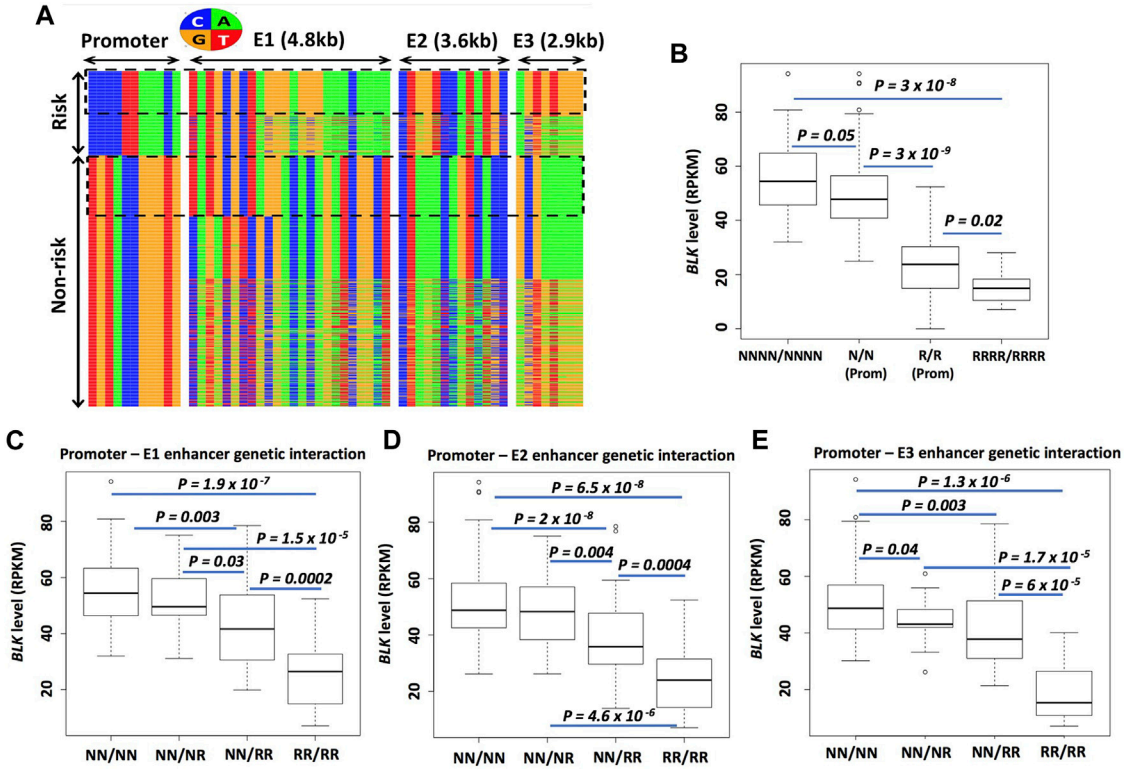


FIGURE 7
Enhancer haplotypes modulate *BLK* expression. **(A)** Promoter-E1-E2-E3 haplotypes conditioned on promoter risk and non-risk haplotypes. Each row represents a haplotype and each column - a SNP. Nucleotides are color-coded as shown in the circle above the haplotype heatmap **(B)** Risk haplotypes at enhancers lower *BLK* expression. **(C–E)** "Risk dosage" effect. The *BLK* expression level tends to decrease as the total number of 'risk' haplotypes in the regulatory regions (promoter and enhancers E1, E2, E3) increases. In **(B–E)**, N means non-risk haplotype and R means risk haplotype. NNNN/NNNN and RRRR/RRRR denote multilocus genotypes at P-E1-E2-E3; N/N and R/R denote multilocus genotypes at the promoter; NN/NN, NN/NR and so on in **(C–E)** denote multilocus genotypes at P-E1, P-E2 or P-E3. See Materials and Methods for details on the notation.

heterozygous SNPs and labeled as one of the following using phased genotype data from the 1000 Genomes Project: parent-1 allele (M), parent-2 allele (P), or allele unassigned (UA). The M-M, M-UA, and UA-M paired-end reads were grouped as "parent-1" and P-P, P-UA and UA-P paired-end reads were grouped as "parent-2" to compute chromatin interaction frequencies on two parental chromosomes at 5 kb resolution. To identify differences in chromatin interaction frequencies of homologous chromosomes, we determined haplotype-specific interaction frequency matrices in 5 kb bins for each replicate library. For n replicate libraries from each cell line, we computed n matrices for parent-1 haplotype and n matrices for parent-2 haplotype. Treating the entries of these matrices as sequence count data with the study design {M, M, . . . , M, P, P, . . . , P} (n consecutive M's followed by n consecutive P's), we determined differences between chromatin interaction frequencies on parent-1 and parent-2 chromosomes using multiHiCcompare (version 1.10.0) method (Stansfield et al., 2019) with the following parameters: make_hicexp (zero.p = 0.8, A.min = 5), cyclic_loess (span = 0.2), logfc_cutoff = 0.5, logcpm_cutoff = 0.5, p. method = "fdr" and p.adj_cutoff = 0.05. Differential chromatin interaction data was uploaded as custom tracks to WashU Epigenome Browser (https://epigenomegateway.wustl.edu/) and the images were then exported (Figures 2,3A).

Enhancer haplotype analysis

Using Bayesian approaches to describe "credible sets" of disease-causal SNPs following published methods (Wellcome Trust Case Control Consortium et al., 2012), we have identified the smallest set of SNPs accounting for 95% of the posterior probability from 3,892 European-American SLE cases and 3,464 controls (Rasmussen et al., 2011). This credible causal set at *FAM167A-BLK* locus, henceforth denoted as BLK-CS-12, consists of 12 SNPs that are in high linkage disequilibrium (LD) ($r^2 > 0.9$), spans a 13 kb region near *BLK* promoter. Conditional analysis on any of the 12 markers reduces association for the other 11, consistent with a single genetic association. Since the 12 markers are in perfect linkage disequilibrium with each other,

without loss of generality, we have omitted a rare indel variant (rs202125301) from BLK-CS-12 and used the set of 11 SNPs (Supplementary Table S5), which we denote as BLK-CS-11, in our analyses leading to Figure 7.

We have written a collection of Perl and R codes to perform analyses leading to Figure 7. We obtained phased genotype VCF file for the European population (GBR, FIN, CEU, IBS, and TSI; 503 individuals in total) from the 1000 Genomes Project (version v5a.20,130,502), and retained genetic variants with minor allele frequency (MAF) ≥ 5% for downstream analysis. From the resulting VCF file, we have extracted phased genotype data for 11 SNPs from the BLK-CS-11 set, 24 SNPs from the enhancer E1 region (hg19 coordinate chr8: 11,387,781–11,392,710), 13 SNPs from enhancer E2 region (chr8:11,393,832–11,397,791) and 8 SNPs from enhancer E3 region (chr8:11,402,260–11,405,263). These SNP numbers correspond to the number of columns (11 + 24+13 + 8) in the heatmap plot in Figure 7A. From the phased genotype data for 11 promoter variants from BLK-CS-11, we extracted 1,006 haplotypes (from 503 individuals), of which 725 are "non-risk" (TGTACCGGGTG), 243 are "risk" (CCCCTTAAACA) and 38 are "idiosyncratic." We omitted the idiosyncratic haplotypes from the downstream analyses. The rows in the heatmap plot in Figure 7A correspond to the haplotypes, with the rows labeled as 'Risk' representing 243 promoter "risk" haplotypes and the rows labeled as "Non-risk" representing 725 promoter "non-risk" haplotypes, respectively. Each entire row in the heatmap represents the concatenation of promoter, E1, E2, and E3 local haplotypes from the same chromosome.

Conditioned on the promoter haplotype, we further stratified haplotypes on the basis of the local haplotype sequences at candidate enhancer E1, E2, and E3 regions as described in detail in the Main text. In Figure 7B, NNNN/NNNN denotes a multilocus genotype at P-E1-E2-E3. It denotes all individuals in the dataset D344 [gene expression and genotype data from 344 European individuals (Lappalainen et al., 2013)] who possess the P-E1-E2-E3 haplotype shown in the lower dotted box in the 'Non-risk' block in Figure 7A on both parental chromosomes 8. Similarly, RRRR/RRRR denotes two identical haplotypes from the upper dotted box in the "Risk" block in Figure 7A.

N/N (prom) in Figure 7A denotes individuals from the D344 dataset who possess prom-N (i.e. non-risk) promoter haplotype (TGTACCGGGTG) on both chromosomes 8, with the local haplotypes at E1, E2, and E3 unspecified. In other words, an individual from the N/N (prom) group can have any promoter-E1-E2-E3 haplotype shown in the "Non-risk" block in Figure 7A. Similarly, an individual from the R/R (prom) group can have any promoter-E1-E2-E3 haplotype shown in the "Risk" block in Figure 7A.

NN/RR in Figure 7E denotes the multilocus genotype (TGTACCGGGTG)-(GCGAAAAA)/(CCCCTTAAACA)-(AGTGTGGG) at promoter-E3, with the local haplotypes at E1 and E2 unspecified. Other multilocus genotypes in Figures 7C–E are similarly defined.

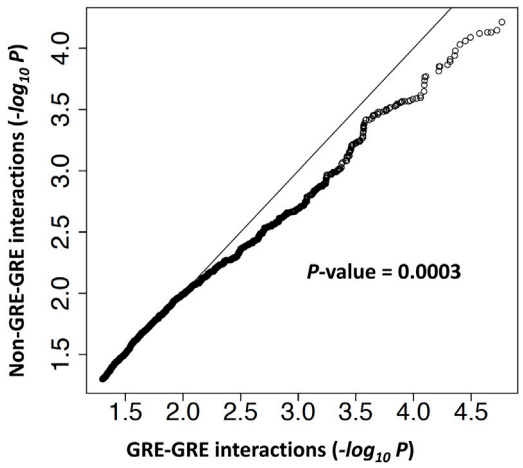


FIGURE 8
Genetic interactions are enriched at gene regulatory elements. A quantile-quantile (Q–Q) plot of -log10 *p*-values from the log-likelihood ratio test of genetic interaction. Genetic interactions were separated into two groups. In the GRE-GRE group, both SNPs are located in gene regulatory regions. In the non-GRE-GRE group, at least one SNP is located outside of the regulatory regions.

The one-sided Student's t-test was used to compare *BLK* expression level distributions in the two groups of individuals in Figures 7B–E.

Statistical test of genetic interactions enrichment in regulatory regions

From the dataset D344, we retrieved SNPs from the 143 kb region R143k (hg19 coordinate chr8:11, 331, 000–11,474,000) at the *BLK* locus. The log-likelihood ratio test was used to determine statistically significant genetic interactions. For each pair of SNPs from the region R143k, we fitted a linear ($y = b_0 + b_1x_1 + b_2x_2$) and an interaction ($y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2$) model to *BLK* expression and genotype data, where y denotes expression level and x_1, x_2 denote 0/1/2 coded genotypes of the two SNPs under consideration. The log-likelihood ratio test was used to determine statistically significant genetic interactions i.e. the test of whether the interaction model explains the observed data better than the linear model does.

For each pair of SNPs from the region R143k, we thus have a *p*-value from the log-likelihood test of genetic interaction. To test the hypothesis that genetic interactions are enriched in gene regulatory regions at the *BLK* locus, we separated SNPs from the R143k region into two groups: SNPs from the

putative regulatory regions of *BLK*, as defined by the H3K27ac epigenetic mark and REST binding region (see Figure 2; Supplementary Table S7) (denoted as GRR) and SNPs outside of the putative regulatory regions. For each SNP pair (SNP1, SNP2), if both SNPs belong to GRR, then the SNP pair is of the "GRE-GRE" type in the notation of Figure 8. If at least one of the two SNPs is outside of GRR, the SNP pair is of the "Non-GRE-GRE" type. The question of whether the genetic interactions are predominantly between SNPs located in gene regulatory regions thus reduces to the problem of comparison of the distribution of the genetic interaction log-likelihood *p*-values defined above for the group of GRE-GRE SNP pairs and non-GRE-GRE SNP pairs. We used the quantile-quantile plot (Q-Q plot) plot to show that the *p*-values of SNP-SNP interactions in the GRE-GRE group are more significant than those in the non-GRE-GRE group. One-sided Kolmogorov-Smirnov test was used to compare the distribution of the log-likelihood *p*-values to obtain *p*-value = 0.0003 shown in Figure 8.

Discussion

Why are our findings on genetic interactions important for SLE? After all, credible causal SNPs are clearly localized to the *BLK* promoter region. Our hypothesis is that SNPs located in enhancer regions can further contribute to disease risk by modulating the effects of promoter SNPs on *BLK* expression. Indeed, a previous model proposes that genetic variants in weak linkage disequilibrium (LD) with risk variants can influence disease risk via physical interactions in the 3D chromatin context (Corradin et al., 2016). Our functional experiments and genetic analyses provide explicit evidence supporting the validity of this model.

In this study, we presented evidence that global haplotype-specific 3D chromatin interactions between regulatory regions can have a strong influence on local allelic gene expression, and consequently, on disease risk. Our capture Hi-C data revealed that many long-range chromatin interactions in the 8p23.1 region are haplotype-specific. Focusing specifically on the *FAM167A-BLK* locus, we found that the *BLK* promoter and the enhancer E3 located 52 kb downstream from the promoter haplotype-specifically interact in the 3D chromatin context, with the interaction being weaker on the SLE risk haplotype, consistent with the reduced expression of *BLK* on the risk haplotype (Figures 2,3). Interestingly, we found that the enhancer E3 interacts with haplotype-specificity with two nearby enhancers E1 and E2, which are members of the same super-enhancer cluster to which E3 belongs (Hnisz et al., 2013), with the interactions being weaker on the risk haplotype. We hypothesized that these interactions amplify the enhancer activity of E3. To test this hypothesis, we have performed allele-specific enhancer reporter assay and allele-specific H3K27ac ChIP-qPCR experiments (Figure 5). Our reporter experiments revealed that enhancer activity of non-risk E3 sequence is ~1.5 times higher than the activity of risk E3 sequence (Figure 5A). Consistent with our 'local enhancer activity amplification due to long-range chromatin interactions' hypothesis, allele-specific ChIP-qPCR experiments demonstrated that H3K27ac signal at E3 enhancer on non-risk haplotype is ~8.5 times higher than on risk haplotype (Figure 5B), the reasoning being that enhancer activity in reporter assays lacks long-range chromatin context whereas the H3K27ac ChIP-qPCR measures endogenous enhancer activity in this particular chromatin context. Together, these findings support our hypothesis that local allele-specific enhancer activities are influenced by global haplotype structure due to chromatin looping interactions.

Our chromatin interaction data also revealed that *BLK* promoter interacts haplotype-specifically with a distal repressor REST binding site and the *FAM167A* gene promoter, with the interactions being stronger on the SLE risk haplotype (Figures 2,3), consistent with reduced expression of *BLK* on the risk haplotype, and the hypothesis that *BLK* promoter is "enhancer-like" and may regulate expression of *FAM167A*, respectively. For recent research on transcriptional regulation by "enhancer-like" promoters, see (Xu et al., 2011; Dao et al., 2017; Dao and Spicuglia, 2018). We tested our hypothesis on the regulation of *FAM167A* expression by the "enhancer-like" *BLK* promoter using dCas9-KRAB chromatin repressor experiments (Figure 4). These experiments have provided strong evidence in support of our hypothesis.

These results nominate *FAM167A*, in addition to *BLK*, or both as potential risk genes at the *FAM167A-BLK* locus. If the regulatory structure we present herein is responsible for the change in disease risk for SLE, and maybe, other disorders associated with the *FAM167A-BLK* locus, rather than a different regulatory structure in another cell type, then perhaps *FAM176A* is the true gene mediating disease risk for two reasons. First, down-regulating signal transduction from the B cell receptor by reducing *BLK* expression would seem to be counter-intuitive for a disease like SLE where autoantibody generation is central to pathogenesis. Second, DIORA-1 (the gene product of *FAM167A*) has an activity that could well have a profound influence on the inflammatory response. DIORA-1 is a disordered protein (Mentlein et al., 2018) that is secreted and binds desmoglein-1 (DSG1) to gain cell entry, which then activates NFκB via its non-canonical pathway by liberating NFκB-inhibitor kinase from DSG1. DIORA-1 appears responsible for much BCR-ABL-tyrosine kinase inhibitor resistance in chronic myelogenous leukemia (CML) (Yang et al., 2022). Certainly, DIORA-1 has functional properties that make the level of activity of this gene product attractive for mediating SLE risk. An understanding of the evolutionary advantage for *BLK* and *FAM167A* to be reciprocally regulated in the way we describe awaits a deeper understanding of the inter-relationships of the pathways impacted by their two gene products. However, the multiple diseases with risk variants at this locus would be consistent with a continuing evolutionary impact.

Haplotype-specific enhancer-enhancer and enhancer-promoter chromatin interactions (Figure 2), and evidence for the "local enhancer activity amplification due to long-range chromatin

interactions" hypothesis (Figure 5) have led us to seek functional evidence of synergistic interactions of *BLK* transcriptional regulatory elements. Using dCas9-p300 CRISPR epigenome editing experiments, we have activated the silent *BLK* locus in K562 cells and demonstrated enhancer-enhancer and enhancer-promoter synergies in *BLK* activation (Figure 6).

Our genetic analyses have revealed that enhancer haplotypes can modulate *BLK* expression (Figure 7) and suggested a "risk dosage" model whereby disease risk alleles at multiple regulatory elements at *BLK* locus synergistically decrease gene expression, and consequently, increase disease risk.

These studies have been done with LCLs, which are generally, B cell lines that have been infected by Epstein-Barr virus (EBV), which is strong etiologic candidate for causing SLE and Multiple Sclerosis (MS) (Harley and James, 2006; Bjornevik et al., 2022; Laurynenka et al., 2022). The LCL is a stable transformed cell line expressing the Latency III program of EBV. The EBV gene product and transcription co-factor, EBNA2, is concentrated at SLE and MS risk loci, including *BLK* (Harley et al., 2018, p. 2; Yin et al., 2021). Recently, Afrasiabi et al. (2022), have extended these observations by showing that *BLK* and *FAM167A* are bound by EBNA2 and the products of both genes are differentially expressed as eQTLs in greater magnitude in LCLs than in B cells that are not EBV infected, with *BLK* and *FAM167A* being affected in opposing directions. Both *BLK* and *FAM167A* are correlated with EBV DNA copy number per cell, with the association with *FAM167A* being much more convincing. Finally, the level of EBNA2 in their data is inversely proportional to the level of *BLK* expression. These observations add another level of complexity relating the environment to disease risk that begs for an understanding of how these differences may or may not be components of mechanisms that influence disease risk.

Common SNPs with effect sizes well below genome-wide statistical significance account for a large proportion of "missing heritability" of many traits (Yang et al., 2010). However, mechanistic details of how weak-effect genetic variants contribute to heritability and disease risk remain largely unknown because we still have very limited knowledge of how these variants percolate through the entire cellular and gene regulatory networks (Boyle et al., 2017). Investigations such as those in (Corradin et al., 2016; Boyle et al., 2017), and the present study represent important steps toward deciphering the mechanistic details of the genotype-phenotype map in disease etiology.

Data availability statement

The high-throughput sequencing data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo) under accession number GSE211246. 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

IC conceived the study. IC, MS-JR, PT, and JH designed the study. MS-JR and PT performed the experiments. IC, PT, MS-JR, and BN analyzed the data. IC and JH drafted the article. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by funding (to IC) from the Lupus Research Alliance, Arthritis National Research Foundation, National Psoriasis Foundation, and Cincinnati Children's Research Foundation, and by the following grants (to JH): NIH grants R01 AI024717, R01 AI148276, U01 AI130830, UM1 AI144292, U54 GM104938, and P30 AR073750 and a USDVA Merit Award I01 BX001834.

Acknowledgments

We would like to thank Cincinnati Children's Hospital Medical Center's DNA Sequencing and Genotyping Core, and Viral Vector Core for the sequencing and lentivirus production services provided for this study. We would also like to thank Isaac Harley for his comments on the manuscript and for bringing recent work on the functional characterization of *FAM167A* gene product DIORA-1 to our attention.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1008582/full#supplementary-material

References

Afrasiabi, A., Keane, J. T., Ong, L. T. C., Alinejad-Rokny, H., Fewings, N. L., Booth, D. R., et al. (2022). Genetic and transcriptomic analyses support a switch to lytic phase in Epstein Barr virus infection as an important driver in developing Systemic Lupus Erythematosus. *J. Autoimmun.* 127, 102781. doi:10.1016/j.jaut.2021.102781

Babu, D., and Fullwood, M. J. (2015). 3D genome organization in health and disease: Emerging opportunities in cancer translational medicine. *Nucleus* 6, 382–393. doi:10.1080/19491034.2015.1106676

Bjornevik, K., Cortese, M., Healy, B. C., Kuhle, J., Mina, M. J., Leng, Y., et al. (2022). Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science* 375, 296–301. doi:10.1126/science.abj8222

Borowiec, M., Liew, C. W., Thompson, R., Boonyasrisawat, W., Hu, J., Mlynarski, W. M., et al. (2009). Mutations at the BLK locus linked to maturity onset diabetes of the young and beta-cell dysfunction. *Proc. Natl. Acad. Sci. U. S. A.* 106, 14460–14465. doi:10.1073/pnas.0906474106

Boyle, E. A., Li, Y. I., and Pritchard, J. K. (2017). An expanded view of complex traits: From polygenic to omnigenic. *Cell* 169, 1177–1186. doi:10.1016/j.cell.2017.05.038

Castillejo-López, C., Delgado-Vega, A. M., Wojcik, J., Kozyrev, S. V., Thavathiru, E., Wu, Y.-Y., et al. (2012). Genetic and physical interaction of the B-cell systemic lupus erythematosus-associated genes BANK1 and BLK. *Ann. Rheum. Dis.* 71, 136–142. doi:10.1136/annrheumdis-2011-200085

Chepelev, I., Wei, G., Wangsa, D., Tang, Q., and Zhao, K. (2012). Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Res.* 22, 490–503. doi:10.1038/cr.2012.15

Corradin, O., Saiakhova, A., Akhtar-Zaidi, B., Myeroff, L., Willis, J., Cowper-Sal lari, R., et al. (2014). Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res.* 24, 1–13. doi:10.1101/gr.164079.113

Corradin, O., Cohen, A. J., Luppino, J. M., Bayles, I. M., Schumacher, F. R., and Scacheri, P. C. (2016). Modeling disease risk through analysis of physical interactions between genetic variants within chromatin regulatory circuitry. *Nat. Genet.* 48, 1313–1320. doi:10.1038/ng.3674

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* 107, 21931–21936. doi:10.1073/pnas.1016071107

Dao, L. T. M., and Spicuglia, S. (2018). Transcriptional regulation by promoters with enhancer function. *Transcription* 9, 307–314. doi:10.1080/21541264.2018.1486150

Dao, L. T. M., Galindo-Albarrán, A. O., Castro-Mondragon, J. A., Andrieu-Soler, C., Medina-Rivera, A., Souaid, C., et al. (2017). Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nat. Genet.* 49, 1073–1081. doi:10.1038/ng.3884

de Wit, E., and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* 26, 11–24. doi:10.1101/gad.179804.111

Delgado-Vega, A. M., Dozmorov, M. G., Quirós, M. B., Wu, Y.-Y., Martínez-García, B., Kozyrev, S. V., et al. (2012). Fine mapping and conditional analysis identify a new mutation in the autoimmunity susceptibility gene BLK that leads to reduced half-life of the BLK protein. *Ann. Rheum. Dis.* 71, 1219–1226. doi:10.1136/annrheumdis-2011-200987

Dryden, N. H., Broome, L. R., Dudbridge, F., Johnson, N., Orr, N., Schoenfelder, S., et al. (2014). Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. *Genome Res.* 24, 1854–1868. doi:10.1101/gr.175034.114

Gourh, P., Agarwal, S. K., Martin, E., Divecha, D., Rueda, B., Bunting, H., et al. (2010). Association of the C8orf13-BLK region with systemic sclerosis in North-American and European populations. *J. Autoimmun.* 34, 155–162. doi:10.1016/j.jaut.2009.08.014

Gusev, A., Bhatia, G., Zaitlen, N., Vilhjalmsson, B. J., Diogo, D., Stahl, E. A., et al. (2013). Quantifying missing heritability at known GWAS loci. *PLoS Genet.* 9, e1003993. doi:10.1371/journal.pgen.1003993

Gusev, A., Lee, S. H., Trynka, G., Finucane, H., Vilhjálmsson, B. J., Xu, H., et al. (2014). Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am. J. Hum. Genet.* 95, 535–552. doi:10.1016/j.ajhg.2014.10.004

Guthridge, J. M., Lu, R., Sun, H., Sun, C., Wiley, G. B., Dominguez, N., et al. (2014). Two functional lupus-associated BLK promoter variants control cell-type- and developmental-stage-specific transcription. *Am. J. Hum. Genet.* 94, 586–598. doi:10.1016/j.ajhg.2014.03.008

Harley, J. B., and James, J. A. (2006). Epstein-Barr virus infection induces lupus autoimmunity. *Bull. NYU Hosp. Jt. Dis.* 64, 45–50.

Harley, J. B., Chen, X., Pujato, M., Miller, D., Maddox, A., Forney, C., et al. (2018). Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. *Nat. Genet.* 50, 699–707. doi:10.1038/s41588-018-0102-3

Hilton, I. B., D'Ippolito, A. M., Vockley, C. M., Thakore, P. I., Crawford, G. E., Reddy, T. E., et al. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* 33, 510–517. doi:10.1038/nbt.3199

Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., et al. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U. S. A.* 106, 9362–9367. doi:10.1073/pnas.0903103106

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., et al. (2013). Super-enhancers in the control of cell identity and disease. *Cell* 155, 934–947. doi:10.1016/j.cell.2013.09.053

Hom, G., Graham, R. R., Modrek, B., Taylor, K. E., Ortmann, W., Garnier, S., et al. (2008). Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N. Engl. J. Med.* 358, 900–909. doi:10.1056/NEJMoa0707865

International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN)Harley, J. B., Alarcón-Riquelme, M. E., Criswell, L. A., Jacob, C. O., Kimberly, R. P., Moser, K. L., et al. (2008). Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. *Nat. Genet.* 40, 204–210. doi:10.1038/ng.81

Ito, I., Kawasaki, A., Ito, S., Hayashi, T., Goto, D., Matsumoto, I., et al. (2009). Replication of the association between the C8orf13-BLK region and systemic lupus erythematosus in a Japanese population. *Arthritis Rheum.* 60, 553–558. doi:10.1002/art.24246

Jäger, R., Migliorini, G., Henrion, M., Kandaswamy, R., Speedy, H. E., Heindl, A., et al. (2015). Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nat. Commun.* 6, 6178. doi:10.1038/ncomms7178

Klann, T. S., Black, J. B., Chellappan, M., Safi, A., Song, L., Hilton, I. B., et al. (2017). CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat. Biotechnol.* 35, 561–568. doi:10.1038/nbt.3853

Krijger, P. H. L., and de Laat, W. (2016). Regulation of disease-associated gene expression in the 3D genome. *Nat. Rev. Mol. Cell Biol.* 17, 771–782. doi:10.1038/nrm.2016.138

Lappalainen, T., Sammeth, M., Friedländer, M. R., 't Hoen, P. A. C., Monlong, J., Rivas, M. A., et al. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 501, 506–511. doi:10.1038/nature12531

Laurynenka, V., Ding, L., Kaufman, K. M., James, J. A., and Harley, J. B. (2022). A high prevalence of anti-EBNA1 heteroantibodies in systemic lupus erythematosus (SLE) supports anti-EBNA1 as an origin for SLE autoantibodies. *Front. Immunol.* 13, 830993. doi:10.3389/fimmu.2022.830993

Lessard, C. J., Li, H., Adrianto, I., Ice, J. A., Rasmussen, A., Grundahl, K. M., Kelly, J. A., et al. (2013). Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjögren's syndrome. *Nat. Genet.* 45, 1284–1292. doi:10.1038/ng.2792

Li, G., Ruan, X., Auerbach, R. K., Sandhu, K. S., Zheng, M., Wang, P., et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 148, 84–98. doi:10.1016/j.cell.2011.12.014

Li, Y., Hu, M., and Shen, Y. (2018). Gene regulation in the 3D genome. *Hum. Mol. Genet.* 27, R228–R233. doi:10.1093/hmg/ddy164

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., et al. (2009). Finding the missing heritability of complex diseases. *Nature* 461, 747–753. doi:10.1038/nature08494

Mentlein, L., Thorlacius, G. E., Meneghel, L., Aqrawi, L. A., Ramírez Sepúlveda, J. I., Grunewald, J., et al. (2018). The rheumatic disease-associated FAM167A-BLK locus encodes DIORA-1, a novel disordered protein expressed highly in bronchial epithelium and alveolar macrophages. *Clin. Exp. Immunol.* 193, 167–177. doi:10.1111/cei.13138

Namjou, B., Ni, Y., Harley, I. T. W., Chepelev, I., Cobb, B., Kottyan, L. C., et al. (2014). The effect of inversion at 8p23 on BLK association with lupus in Caucasian population. *PloS One* 9, e115614. doi:10.1371/journal.pone.0115614

Nashi, E., Wang, Y., and Diamond, B. (2010). The role of B cells in lupus pathogenesis. *Int. J. Biochem. Cell Biol.* 42, 543–550. doi:10.1016/j.biocel.2009.10.011

Ooi, L., and Wood, I. C. (2007). Chromatin crosstalk in development and disease: lessons from REST. *Nat. Rev. Genet.* 8, 544–554. doi:10.1038/nrg2100

Perez-Pinera, P., Kocak, D. D., Vockley, C. M., Adler, A. F., Kabadi, A. M., Polstein, L. R., et al. (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* 10, 973–976. doi:10.1038/nmeth.2600

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680. doi:10.1016/j.cell.2014.11.021

Rasmussen, A., Sevier, S., Kelly, J. A., Glenn, S. B., Aberle, T., Cooney, C. M., et al. (2011). The lupus family registry and repository. *Rheumatology* 50, 47–59. doi:10.1093/rheumatology/keq302

Salm, M. P. A., Horswell, S. D., Hutchison, C. E., Speedy, H. E., Yang, X., Liang, L., et al. (2012). The origin, global distribution, and functional impact of the human 8p23 inversion polymorphism. *Genome Res.* 22, 1144–1153. doi:10.1101/gr.126037.111

Sánchez, E., Comeau, M. E., Freedman, B. I., Kelly, J. A., Kaufman, K. M., Langefeld, C. D., et al. (2011). Identification of novel genetic susceptibility loci in African American lupus patients in a candidate gene association study. *Arthritis Rheum.* 63, 3493–3501. doi:10.1002/art.30563,

Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C.-J., Vert, J.-P., et al. (2015). HiC-pro: an optimized and flexible pipeline for hi-C data processing. *Genome Biol.* 16, 259. doi:10.1186/s13059-015-0831-x

Simpfendorfer, K. R., Olsson, L. M., Manjarrez Orduño, N., Khalili, H., Simeone, A. M., Katz, M. S., et al. (2012). The autoimmunity-associated BLK haplotype exhibits cis-regulatory effects on mRNA and protein expression that are prominently observed in B cells early in development. *Hum. Mol. Genet.* 21, 3918–3925. doi:10.1093/hmg/dds220

Stansfield, J. C., Cresswell, K. G., and Dozmorov, M. G. (2019). multiHiCcompare: joint normalization and comparative analysis of complex Hi-C experiments. *Bioinformatics* 35, 2916–2923. doi:10.1093/bioinformatics/btz048

Suarez-Gestal, M., Calaza, M., Endreffy, E., Pullmann, R., Ordi-Ros, J., Sebastiani, G. D., et al. (2009). Replication of recently identified systemic lupus erythematosus genetic associations: a case-control study. *Arthritis Res. Ther.* 11, R69. doi:10.1186/ar2698

Thakore, P. I., D'Ippolito, A. M., Song, L., Safi, A., Shivakumar, N. K., Kabadi, A. M., et al. (2015). Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat. Methods* 12, 1143–1149. doi:10.1038/nmeth.3630

Thalayasingam, N., Nair, N., Skelton, A. J., Massey, J., Anderson, A. E., Clark, A. D., et al. (2018). CD4+ and B Lymphocyte expression quantitative traits at rheumatoid arthritis risk loci in patients with untreated early arthritis: Implications for causal gene identification. *Arthritis Rheumatol.* 70, 361–370. doi:10.1002/art.40393

Tsuchiya, N., Ito, I., and Kawasaki, A. (2010). Association of IRF5, STAT4 and BLK with systemic lupus erythematosus and other rheumatic diseases. *Nihon Rinsho Meneki Gakkai Kaishi* 33, 57–65. doi:10.2177/jsci.33.57

Wellcome Trust Case Control ConsortiumMaller, J. B., McVean, G., Byrnes, J., Vukcevic, D., Palin, K., Su, Z., et al. (2012). Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* 44, 1294–1301. doi:10.1038/ng.2435

Xu, Z., Wei, G., Chepelev, I., Zhao, K., and Felsenfeld, G. (2011). Mapping of INS promoter interactions reveals its role in long-range regulation of SYT8 transcription. *Nat. Struct. Mol. Biol.* 18, 372–378. doi:10.1038/nsmb.1993

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42, 565–569. doi:10.1038/ng.608

Yang, T., Sim, K.-Y., Ko, G.-H., Ahn, J.-S., Kim, H.-J., and Park, S.-G. (2022). FAM167A is a key molecule to induce BCR-ABL-independent TKI resistance in CML via noncanonical NF-κB signaling activation. *J. Exp. Clin. Cancer Res.* 41, 82. doi:10.1186/s13046-022-02298-1

Yin, H., Borghi, M. O., Delgado-Vega, A. M., Tincani, A., Meroni, P.-L., and Alarcón-Riquelme, M. E. (2009). Association of STAT4 and BLK, but not BANK1 or IRF5, with primary antiphospholipid syndrome. *Arthritis Rheum.* 60, 2468–2471. doi:10.1002/art.24701

Yin, X., Kim, K., Suetsugu, H., Bang, S.-Y., Wen, L., Koido, M., et al. (2021). Meta-analysis of 208370 East Asians identifies 113 susceptibility loci for systemic lupus erythematosus. *Ann. Rheum. Dis.* 80, 632–640. doi:10.1136/annrheumdis-2020-219209

Zuk, O., Hechter, E., Sunyaev, S. R., and Lander, E. S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci. U. S. A.* 109, 1193–1198. doi:10.1073/pnas.1119675109

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 28 September 2022
DOI 10.3389/fgene.2022.991314

Check for updates

OPEN ACCESS

EDITED BY
Yonghu Sun,
Shandong Provincial Hospital of Dermatology, China

REVIEWED BY
Likui Feng,
The Rockefeller University, United States
Haiyun Li,
Xi'an Jiaotong University, China
Jinsong Wang,
Hubei Normal University, China

*CORRESPONDENCE
Guannan Bai,
guannanbai@zju.edu.cn
Lidan Hu,
hulidan@zju.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

RECEIVED 11 July 2022
ACCEPTED 29 August 2022
PUBLISHED 28 September 2022

CITATION
Xu J, Zhang Y, Zhu K, Li J, Guan Y, He X, Jin X, Bai G and Hu L (2022), Clinical characteristics and in silico analysis of congenital pseudarthrosis of the tibia combined with neurofibromatosis type 1 caused by a novel NF1 mutation. *Front. Genet.* 13:991314. doi: 10.3389/fgene.2022.991314

COPYRIGHT
© 2022 Xu, Zhang, Zhu, Li, Guan, He, Jin, Bai and Hu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Clinical characteristics and in silico analysis of congenital pseudarthrosis of the tibia combined with neurofibromatosis type 1 caused by a novel *NF1* mutation

Jingfang Xu[1†], Ying Zhang[2†], Kun Zhu[3], Jiabin Li[4], Yuelin Guan[5], Xinyu He[5], Xuejing Jin[6], Guannan Bai[7*] and Lidan Hu[5*]

[1]Department of Orthopaedics, The Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, China, [2]Institute of Translational Medicine, Zhejiang University School of Medicine, Hangzhou, China, [3]Department of Pathology, The Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, China, [4]Department of Pharmacy, The Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, China, [5]The Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, China, [6]Centre for Evidence-based Chinese Medicine, Beijing University of Chinese Medicine, Beijing, China, [7]Department of Child Health Care, The Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, China

Congenital pseudarthrosis of the tibia (CPT) is a rare congenital bone malformation, which has a strong relationship with Neurofibromatosis type 1 (NF1). NF1 is an autosomal dominant disease leading to multisystem disorders. Here, we presented the genotypic and phenotypic characteristics of one unique case of a five-generation Chinese family. The proband was CPT accompanied with NF1 due to *NF1* mutation. The proband developed severe early-onset CPT combined with NF1 after birth. Appearance photos and X-ray images of the left limb of the proband showed significant bone malformation. Slit-lamp examination showed Lisch nodules in both eyes of the proband. Whole-exome sequencing (WES) and Sanger sequencing confirmed the truncation variant of *NF1* (c.871G>T, p. E291*). Sequence conservative and evolutionary conservation analysis indicated that the novel mutation (p.E291*) was highly conserved. The truncated mutation led to the loss of functional domains, including CSRD, GRD, TBD, SEC14-PH, CTD, and NLS. It may explain why the mutation led to a severe clinical feature. Our report expands the genotypic spectrum of NF1 mutations and the phenotypic spectrum of CPT combined with NF1.

KEYWORDS

neurofibromatosis type 1, congenital pseudarthrosis of the tibia, ferroptosis, bone deformity, genetic mutation

Introduction

Congenital pseudarthrosis of the tibia (CPT) is a rare type of bone deformity with typical characteristics of pseudarthrosis in early life, pathological fractures of the anterolateral part of the tibia that result in bowing, narrowing of the medullary canal, or a cyst (Crawford, 1986; Hefti et al., 2000). The prevalence is approximately one in 140,000–250,000 births (Crawford and Schorry, 1999; Kesireddy et al., 2018). Previous experience and research suggested that there was a strong relationship between CPT and neurofibromatosis type 1 (NF1) (Crawford, 1986), but the prevalence was heterogeneous worldwide (Hefti et al., 2000; Vander Have et al., 2008). A recent review by van Royen et al. (2016) reported that the prevalence of NF1 in patients with CPT was around 84.0%.

NF1 is caused by mutations in the *NF1* tumor suppressor gene on the 17q11.2 chromosome and affects multiple systems including neurocutaneous and skeletal systems characterized by various clinical manifestations including the typical Café-au-lait macules, Lisch nodules, bone deformity, and multiple neurofibromas (Young et al., 2002; Jett and Friedman, 2010). Treatment for CPT is very difficult and challenging, which is often referred to as one of the most puzzling and frustrating conditions in pediatric orthopedics worldwide (Granchi et al., 2012). Patients may need multiple surgeries frequently to consolidate the pseudarthrosis and unfortunately, refractures occur. After the surgery, there is still a risk of not achieving bone union, and the risk of amputation will be never completely avoided. CPT complicated with NF1 makes the treatment even more challenging, which significantly impairs patients' growth and development, mental health, and quality of life, which imposes a heavy burden on children, families, and society. The etiology of CPT has not been completely understood and little is known about why CPT is closely related to NF1. Hereditary factors play an essential role in pathogenesis.

The tumor suppressor gene *NF1* comprises 350 kb of genomic DNA (Viskochil et al., 1990) and contains 60 exons encoding a large neurofibromin protein of 2,818 amino acids (Marchuk et al., 1991). To date, pathogenic mutations in the *NF1* gene have been detected as point mutations, frame deletions or duplications, indels, and complex rearrangements (Human Gene Mutation Database (HGMD, http://www.hgmd.org/). Most of these variations result in the truncated protein product. Thus, the NF1 deficiency or the loss of neurofibromin function is related to the increased Ras activity and a high proliferation rate (Trovó-Marqui and Tajara, 2006). The development of Whole-exome sequencing (WES) makes the diagnosis and potential targeting of gene therapy possible. Here, we reported a novel truncated variant (c.871G>T; p. E291*) in a Chinese pedigree and elucidated the possible functional loss through bioinformatic analysis.

Materials and methods

Proband, pedigree, and clinical assessments

This study was conducted according to the principles of the Declaration of World Medical Association (2013). It has been approved by the Local Research Ethics Committee of the Children's Hospital, Zhejiang University School of Medicine (2022-IRB-148). In March 2022, a boy of 14 years and 11 months visited the Department of Orthopedics in the Children's Hospital, Zhejiang University School of Medicine, Hangzhou, China. The main complaints were deformities of the left ankle and the left limb after birth and the inability to walk due to the severe pain in the left limb for 2 weeks. He was considered as the proband in the present study. In addition, he had scoliosis of the spine and multiple café-au-lait macules all over the body except for the face. Comprehensive examinations and tests were conducted, including X-ray examinations of limbs and spine, three-dimensional computed tomography (CT) of the left tibia and fibula, visual examinations and slit-lamp examination, blood routine examination, liver function, renal function, routine urine test, routine stool test, bone metabolism test, Vitamin D test, blood clotting function test, and detections for Human Immunodeficiency Virus (HIV), Syphilis, Hepatitis B, and Hepatitis C. The family history of NF1 and CPT was asked by the orthopedist and information on the five-generation Chinese family was collected. Based on the above information, the diagnosis of the proband was made as CPT combined with NF1. The proband underwent tibial osteotomy and correction and intramedullary nail fixation. At present, a small amount of callus growth can be seen. His mother and grandmother (i.e., his mother's mother) were observed with café-au-lait macules, dermatofibroma, freckles, and severe scoliosis and kyphosis (only seen in the mother) and a diagnosis of NF1 was made according to the establishing criteria. Grandfather (i.e., mother's father), father, and younger brother of the proband were also included in the present study. They had a relatively good health status without any CPT and NF1-related symptoms or signs.

Whole-exome sequencing

Genomic DNA was extracted from peripheral leukocytes using the Blood Genome Qiagen Blood DNA mini kit (Qiagen). The whole-exome DNA library was prepared by using VAHTS Universal DNA Library Prep Kit for Illumina V3 (Vazyme) and KAPA HyperExome Sequence Capture Kit (Roche, United States). The Illumina DNA Standa ds and Primer Premix Kit (kapa) was used for library quantification, followed by sequencing with the dnBSEQ-T7 gene sequencer (PE150). Raw data with low-quality reads were filtered for quality control. The Burrows-Wheeler Aligner (BWA) sequence alignment method was performed by comparing with the human genome reference (hg19). The mutation sites in the target sequence were identified using GATK software. Familial

TABLE 1 Basic information, clinical manifestations, laboratory tests and image tests of the three affected members.

Characteristics	V-1	IV-2	III-4
Age (years)	14.9	35	55
Onset age (years)	At birth	13	/
Sex	Male	Female	Female
Height (cm)	161	145	155
Weight (kg)	43	40	38
Clinical manifestations	1. Bone deformity	1. Café-au-lait macules	1. Café-au-lait macules
	2. Café-au-lait macules	2. Dermatofibroma	2. Dermatofibroma
	3. Axillary and inguinal freckles	3. Axillary and inguinal freckles	3. Axillary and inguinal freckles
	4. Scoliosis of spine	4. Scoliosis and kyphosis	/
Laboratory tests	1. Vitamine D: 2.91 (20–80 ng/ml)	/	
	2. β-collagen specific sequence: 2.292 (0.196–1.665 ng/ml); 25-HydroxyvitaMin D3: 38.5 (41.7–175.0 nmol/L); Bone alkaline phosphatase quality: 21.7 (23.4–119.1 ng/ml)		
Image tests	1. Pseudarthrosis, ankle valgus and foot deformity	Scoliosis and kyphosis	/
	2. Unequal length of both lower limbs		
	3. Scoliosis of spine		
	4. Loss of bone mass		
	5. Lisch nodules		

segregation analysis of identified mutations was carried out whenever applicable. The variants were validated by Sanger sequencing analysis by using TaKaRa LA PCR™ Kit Ver.2.1 (TaKaRa) with ABI 3500XL (Applied Biosystems) platform. The primers were as follows: TGTAAAACGACGGCCAGTTTGCCC TTGGGTTTTTACATAG (forward) and CAGGAAACAGCTATG ACCCCATCAAACAAAGAAACCTAAAATGA (reverse).

Mutation bioinformatics analysis

The American College of Medical Genetics and Genomics (ACMG) Standards and Guidelines for the interpretation of sequence variants were followed in this study (Kalia et al., 2017). The homodimer structure of neurofibromin in open conformation was downloaded from the Protein Data Bank (PDB ID: 7R04). The Pymol visualization tool (http://www.pymol.org/) was used to generate the monomer of NF1. The sequence alignments were analyzed by examining multiple sequence alignments using the program Consurf server and Unipro UGENE.

Results

Clinical information

Table 1 presents the clinical manifestations of the three affected members. Regarding the proband (V-1), the onset age of the disease was at birth. He had bone deformities including the unequal length of lower limbs, bending of left tibia, pseudarthrosis, ankle valgus, and scoliosis of the spine. Multiple café-au-lait macules were distributed all over the body except for the face. Freckles were observed in the axilla and groin. The laboratory tests indicated a low concentration of Vitamin D in the blood and abnormal bone metabolism. The image examinations showed bone deformities, loss of bone mass, and Lisch nodules in both eyes. The mother (IV-2) and grandmother (III-4) had café-au-lait macules distributed all over the body, dermatofibroma around the mouth, on the neck, and in the front chest as well freckles in the axilla and groin. The mother of the proband had severe scoliosis and kyphosis. Figures 1–4 shows the selected results of image examinations, appearance photos, and histopathological findings. Figure 1 shows the presence of pseudarthrosis in the left tibia, valgus of the left ankle, and unequal lengths of the lower limbs. Figure 2 are photos taken under the slit-lamp showing more Lisch nodules in the right eye than that in the left eye. Figures 3A–E shows severe scoliosis and kyphosis of the proband's mother as well as the café-au-lait macules and dermatofibroma on the neck and in the front chest. Figure 3F shows the dermatofibroma distributed around the mouth of the proband's grandmother. Figure 4 are histopathological findings that provided supplementary information to make a diagnosis of neurofibroma.

Genetic characterization and bioinformatic analysis

There are 21 members in this five-generation Chinese family and six individuals were recruited for further study, including

10.3389/fgene.2022.991314

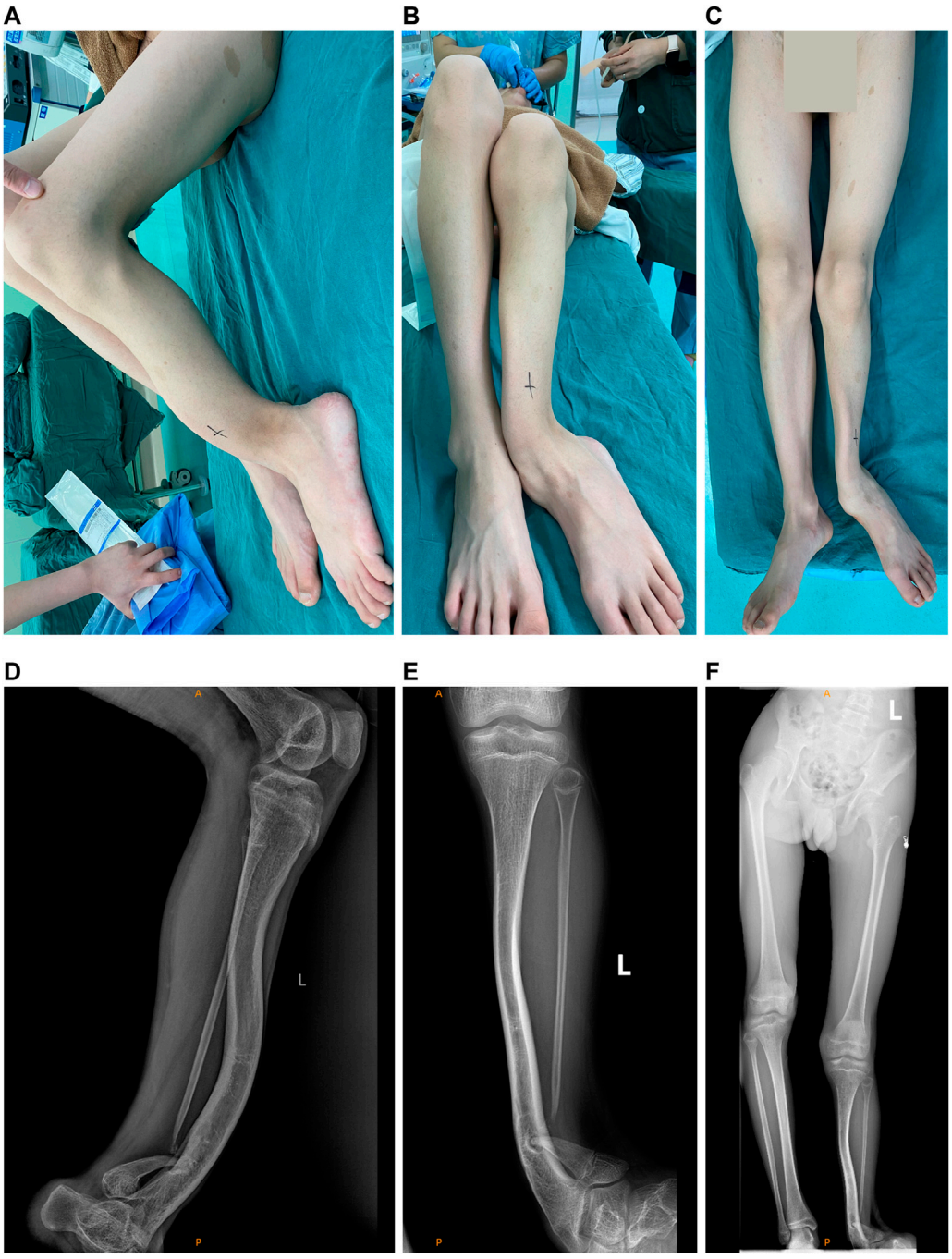


FIGURE 1
Appearance photos and X-ray images of the left limb of the proband. **(A)** Lateral photo of the pseudarthrosis of left tibia. **(B)** Photo of pseudarthrosis of the left tibia and the left ankle valgus. **(C)** Photo of the unequal lengths of the left and right limb. **(D)** X-ray image of the pseudarthrosis of left tibia. **(E)** X-ray image of the pseudarthrosis of the left tibia and the left ankle valgus. **(F)** X-ray image of the unequal lengths of the left and right limb.

three affected individuals and three unaffected individuals. As shown in Figure 5A, the pedigree of the CPT combined with the NF1 patient's family was consistent with autosomal dominant inheritance. To validate the diagnosis, the proband and his brother, parents, and grandparents (i.e., mother's parents) underwent the WES. The mean read depth reached 20x for 98.9% of the target sequences. This was further confirmed by Sanger sequencing (Figure 5B). The genetic test revealed a

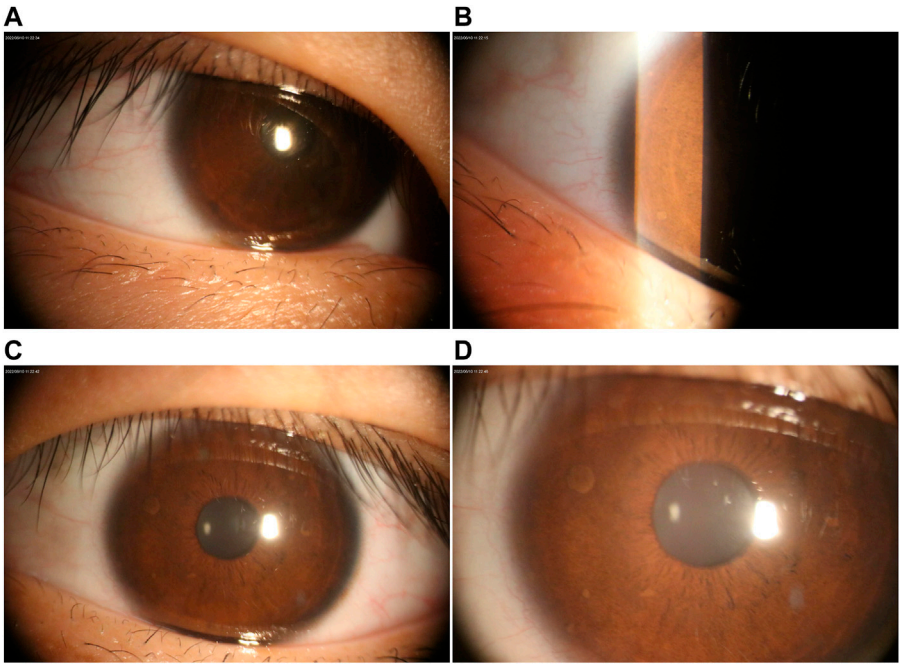


FIGURE 2
Ophthalmological examination under the slit lamp. **(A–B)** Lisch nodules in the left eye. **(C–D)** Right eye with multiple Lisch nodules.

heterozygous truncation variant (c.871G>T, p. E291*) of the exon eight in the *NF1* gene (#OMIM: 162,200). The novel mutation was evaluated as a pathogenic mutation according to ACMG guidelines (2017) (Table 2).

The novel truncation mutation is segregated from the phenotype within the pedigree. This truncation mutation was located near the C-terminal of NF1 and caused the loss of most of the structure of the protein (Figure 6A). The structure of NF1 was displayed in the cartoon and the mutation site (p.E291*) was represented in the magenta sphere (Figure 6B). The cysteine-serine-rich domain (CSRD), GAP-related domain (GRD) and Sec14-homologous domain, and pleckstrin homology domain (SEC14-PH) were rendered light pink, light blue, and pale green, respectively. Sequence conservative analysis indicated that the highly conserved loci of the novel mutation (p.E291*) may be involved in essential physiological functions (Figure 7).

Discussion

CPT is a rare disease and the epidemiological data are limited. The estimated prevalence is around one in 140,000–250,000 births. Due to the different classification systems worldwide, the prevalence or the incidence of CPT is heterogeneous across different countries. In Denmark, Andersen et al. reported a CPT incidence of 1:190,000 live births (Andersen, 1971). In Norway, the incidence of CPT was 1: 60,000 when only taking into account the ethnic Norwegian patients and 1:53,000 when all patients born with CPT in Norway were included (Horn et al., 2013). A strong relationship between CPT and NF1 was suggested. A wide variation in NF1 prevalence numbers, ranging from 0.0 to 84.0%, was demonstrated by the literature review due to the various diagnosis standards (van Royen et al., 2016). Based on the NIH diagnostic criteria that have been considered the golden standard for diagnosis of NF1, the prevalence numbers of NF1 among CPT patients were reported as 42.9%, 54.7%, and 84.0% (Heikkinen et al., 1999; Hefti et al., 2000; van Royen et al., 2016). Previous studies suggested a genetic origin for the etiology of CPT combined with NF1. In this study, the symptoms of the proband were considered to be related to genetic factors because the onset age was very early, i.e., at birth, and it presented with severe symptoms. The pedigree of the CPT NF1 patient's family indicated the pattern of autosomal dominant inheritance and Sanger sequencing confirmed the mutation of *NF1*.

CPT is one of the most puzzling pediatric orthopaedical diseases and the treatment is very challenging. Most of the patients have bending tibia within one or 2 years after birth, and they can also have fractures at birth. Tibial deformities and multiple fractures will eventually form pseudarthrosis, resulting in life-long disability. Current treatment for CPT includes conservative and surgical treatment. Before the child could walk, the physician installs a plaster bracket or a plaster tube to fix the ankle and leg. After the child could walk, the ankle and

10.3389/fgene.2022.991314

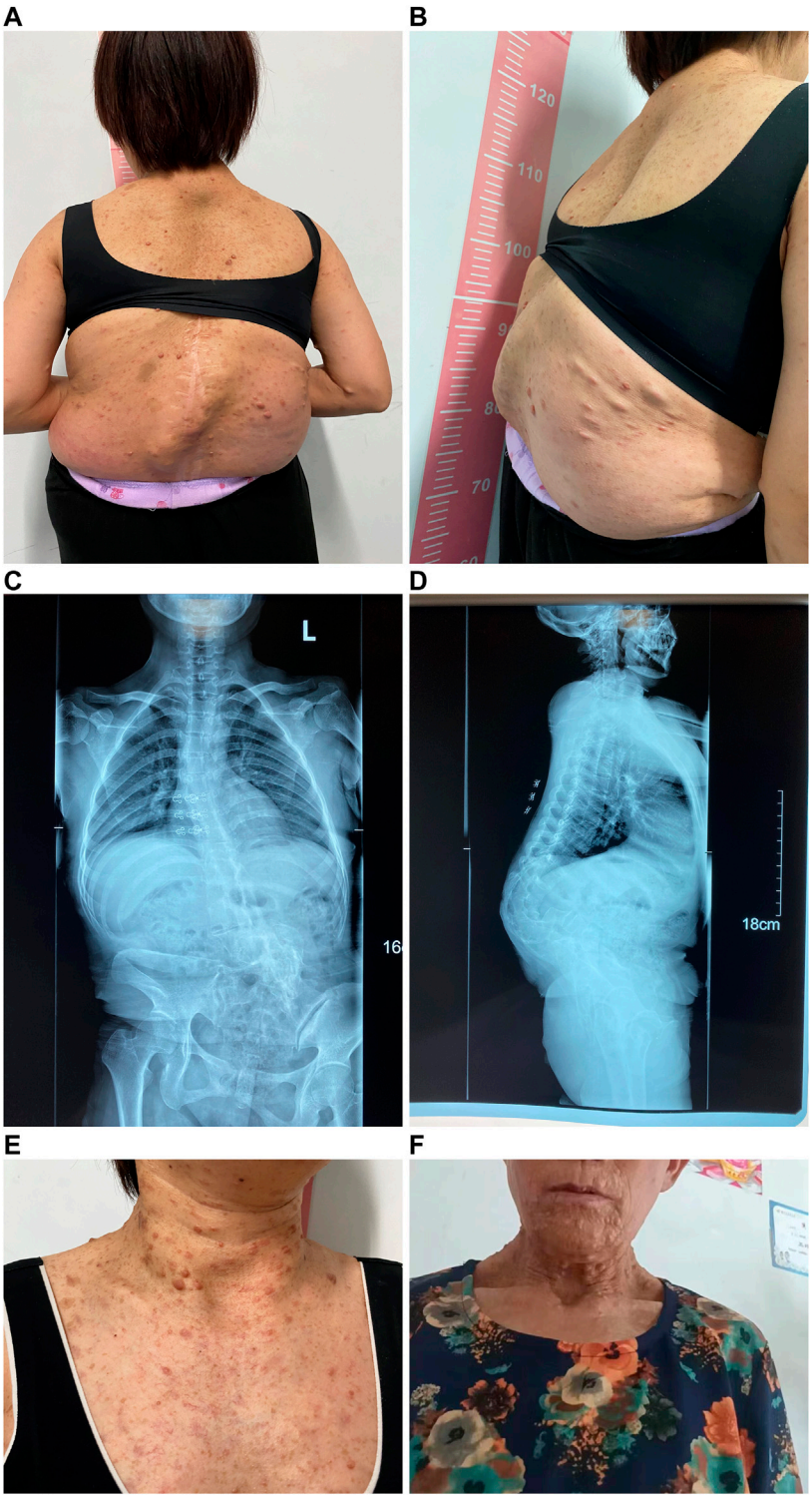


FIGURE 3
Appearance photos and X-ray images of proband's mother. **(A)** Photo of the back with multiple café-au-lait macules, dermatofibroma, and severe scoliosis and kyphosis. **(B)** Lateral photo of the back showing severe scoliosis and kyphosis. **(C)** X-ray image of spinal scoliosis. **(D)** X-ray image of spinal scoliosis and kyphosis. **(E)** Appearance photo of café-au-lait macules and dermatofibroma on the neck and front chest of the proband's mother. **(F)** Appearance photo of dermatofibroma around the mouth of the proband's grandmother.

10.3389/fgene.2022.991314

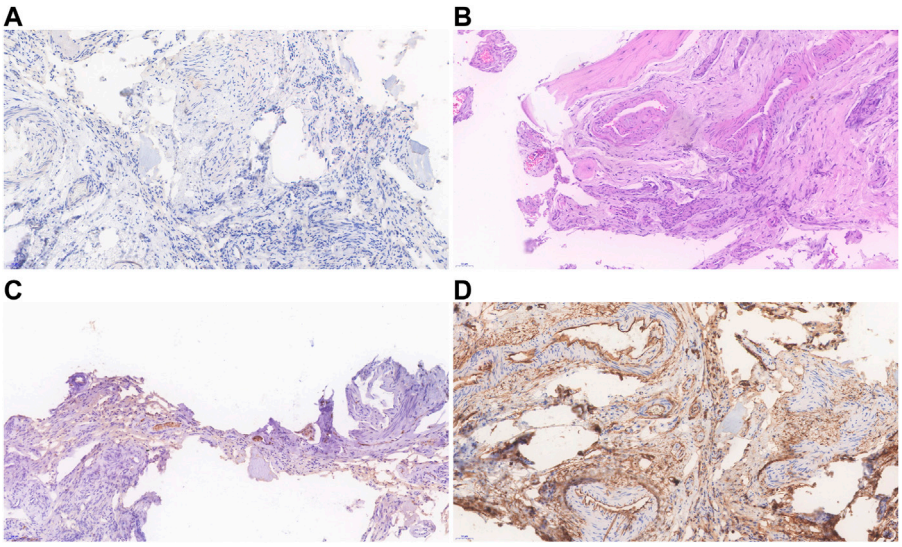


FIGURE 4
Histopathological findings of the proband's neurofibroma. **(A)** Immunohistochemistry showed that the lymphatic marker D2-40 was negative. **(B)** Spindle tumor cells are distributed in bundles with bone destruction, and the tumor is rich in blood vessels. **(C)** Expression of S-100 was scattered in tumor. **(D)** CD34 expression was positive in vascular endothelial cells.

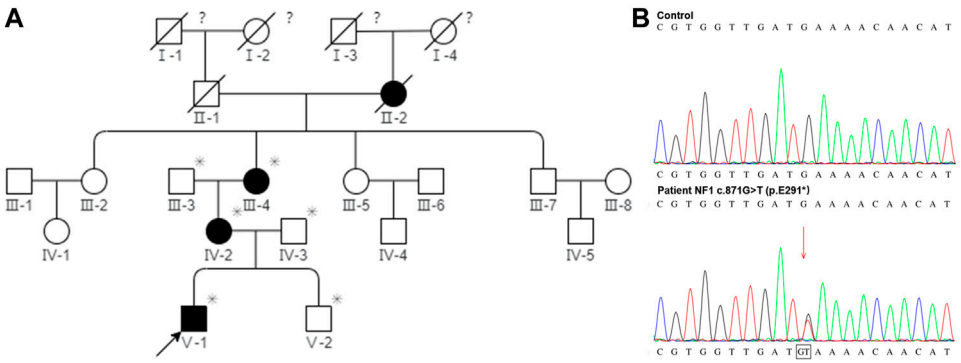


FIGURE 5
Pedigree of the CPT NF1 patient's family and Sanger sequencing confirmed the mutation of NF1. **(A)** Segregation analysis of NF1 variants in a pedigree represents an autosomal dominant inheritance pattern. Squares and circles represent males and females, respectively. Closed and open symbols indicate affected members and unaffected subjects. The arrow denotes the proband. Family history was negative for consanguinity. The slash symbol indicates that the subject is deceased. The asterisk shows the individual underwent both clinical and genetic analyses. The question mark indicates that the ophthalmic history is not available. **(B)** Sanger sequencing at the mutation site with a blood sample. High-throughput sequencing showed a novel *de novo* truncation mutation of the *NF1* gene in the proband was identified (c.871G>T; p. E291*).

TABLE 2 Variant table for the novel mutation of *NF1* gene.

Gene	Chromosome	Exon	Nucleotide change	Amino acid change	ACMG
NF1	Chr17:29509666	8	c.871G>T	p.E291*	Pathogenetic

ACMG: the american college of medical genetics and genomics.

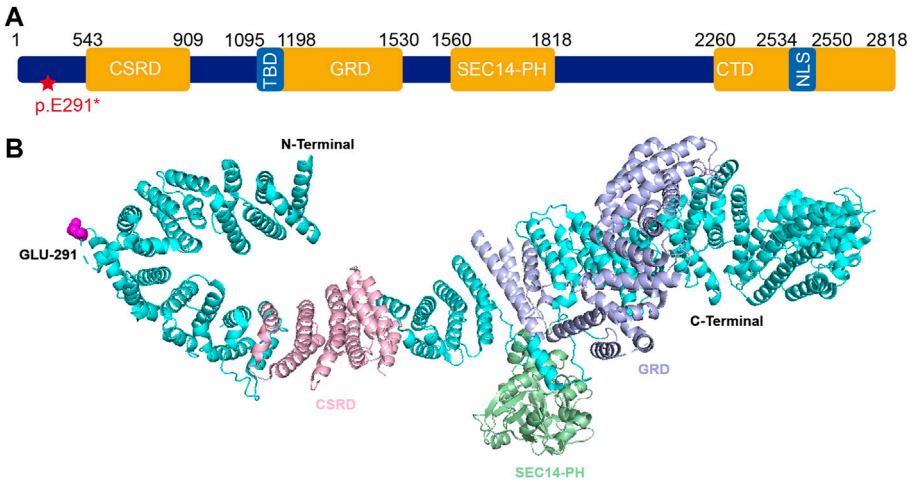


FIGURE 6
(A) Schematic representation of neurofibromin protein and identified mutation of *NF1* (c.871G>T; p. E291*) in this study. **(B)** The 3D structure of neurofibromin1 monomer. The NF1 was represented in cartoon and the mutation site (p.E291*) was shown as magenta sphere. The CSRD domain (residues 543–909) (light pink), GRD domain (residues 1095–1530) (light blue), SEC14-PH domain (residues 1560–1818) (pale green). Abbreviations usedare as follows CSRD, cysteine-serine-rich domain; GRD, GAP-related domain; TBD, tubulin-binding domain; SEC14-PH, Sec14-homologous domain and pleckstrin homology domain; CTD, C-terminal domain; NLS, nuclear localization signal.

leg should be protected with lighter cast support, which may slow down the progression of CPT and avoid the formation of pseudarthrosis and the occurrence of fractures. Surgical treatment can be applied when the child gets older including pseudarthrosis tissue resection, intramedullary rod fixation, wrapped autologous iliac bone transplantation, and Ilizarov annular external fixator compression fixation. Surgery is very challenging and complicating with relatively high risks of nonunion bones, which imposes heavy burdens on the quality of life, psychological well-being, and financial situations of the patients and their families. Mladenov et al. (2020) have reported their treatment protocol for pediatric children with NF1 and tibial pseudarthrosis. Fracture union in tibial pseudarthrosis with satisfactory functional results can be achieved in more than 80% of the children (Mladenov et al., 2020).

The WES could provide further insights into information related to the multisystem disorder. Of note, most of the reported NF1-related variants in the HGMD database are micro changes, and the majority of them are missense/nonsense and small deletions (Figure 8). Three previous studies have investigated the influence of the type of constitutional NF1 mutation on the disease phenotypic variability (Castle et al., 2003; de Luca et al., 2004; Sabbagh et al., 2013). Whereas, limited samples in the survey of Castle et al. and De Luca et al. conferred an ambiguous relationship. Sabbagh et al. revealed a fortuitous association in 565 unrelated patients from the NF-France Network. Understandably, patients with large deletions of the NF1 gene region led to a more severe phenotype. Furthermore, the tendency for truncating mutations to be associated with a greater incidence of Lisch nodules and a larger number of CAL spots as compared with missense mutations. However, these studies failed to find any statistically significant association of NF1 clinical features with mutation type. A more related survey was urgent to clarify the relationship between different genetic and clinical phenotypes.

In the present study, we found a novel truncation mutation (c.871G>T; p. E291*) in a pedigree with a family history of NF1 disorder. The clinical symptoms vary widely among individuals carrying *NF1* gene mutations. The highly conserved protein consists of a cysteine-serine-rich domain (CSRD), a GAP-related domain (GRD), a tubulin-binding domain (TBD), a Sec14-homologous domain, a pleckstrin homology domain (SEC14-PH) and a C-terminal domain (CTD) (Figure 6A). The pathogenic mechanism of CPT is not clear yet and the deletion or inactivation of the NF1 gene is considered as an important cause (Xu et al., 1990; Basu et al., 1992; Cichowski and Jacks, 2001; Viskochil, 2002; Brekelmans et al., 2019). NF1 gene is a tumor suppressor gene with a high mutation rate and extreme mutation heterogeneity. It encodes neurofibroma protein that enhances the GTPase activity of the RAS protein and negatively regulates the Ras/MAPK signaling pathway (Philpott et al., 2017). NF1 gene mutation causes the loss of function of neurofibroma protein, which may enhance the Ras activity, and then result in abnormal proliferation and differentiation of cell lines such as osteoclast progenitor cells and osteoblast progenitor cells (Sharma et al., 2013). The activated MAPK causes osteoblast differentiation, bone formation disorder, and the increased maturation of osteoclast cells, which also explains why fractures of CPT NF1 patients are difficult to heal (Kuorilehto

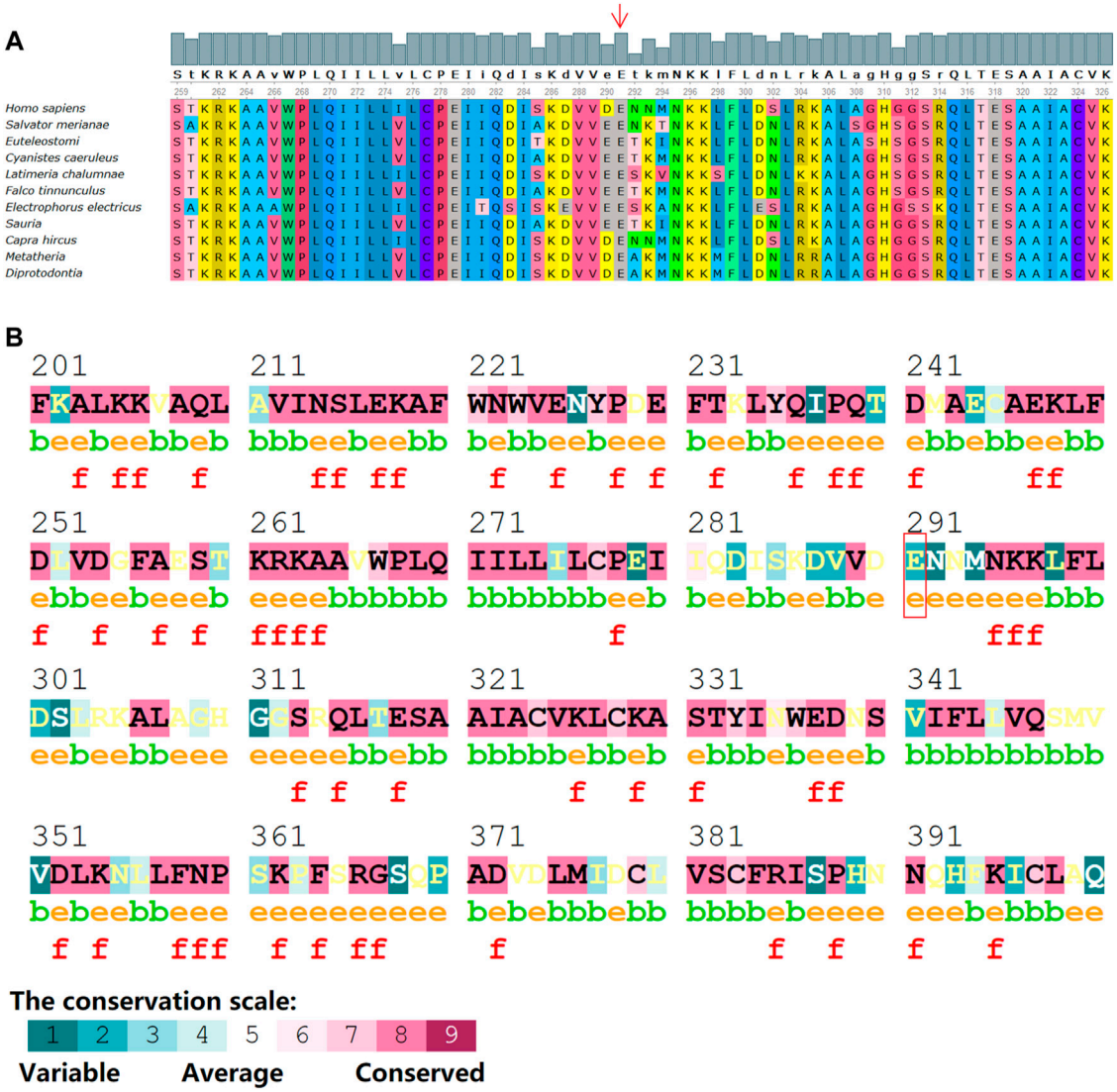


FIGURE 7
Sequence alignments of NF1 protein from various species with program Consurf server and Unipro UGENE software. **(A)** Conserved amino acids background colors are the same among species. Numbers indicate the positions of the amino acid sequences. **(B)** The mutation site (p.E291*) within the conservative amino acid region. The conservation scale ranges from variable (blue), to average (white) to conserved (red). Annotations: e: An exposed residue according to the neural-network algorithm. b: A buried residue according to the neural-network algorithm. f: A predicted functional residue (highly conserved and exposed).

et al., 2006). In other words, the fracture healing normally requires the expression of the NF1 gene that inhibits the activation of the Ras/MAPK pathway. GRD activates the activity of RasGTPase. It is well known that the multiple RAS downstream effectors including the PI3K, ERK, and RalA are highly expressed in NF1 patients. Thus, the dysfunction of GRD leads to neurofibroma tumor genesis and progression. The CSRD also regulates the GRD function *via* its phosphorylation by both kinase A (PKA) and protein kinase C (PKC) (Izawa et al., 1996; Mangoura et al., 2006). The CTD is also regulated by PKA and then regulates the Ras-GAP activity negatively (Tokuo et al., 2001; Feng et al., 2004). The functional domain SEC14-PH within the C-terminal region might be implicated in protein and lipid trafficking (Mousley et al., 2007).

The novel truncation mutation (c.871G>T; p. E291*) has not been reported by other literature. Besides our finding, an early study by Zhu et al. (2019) has identified 25 novel variants among 44 NF1 CPT patients in China. On exon 8, there were two novel mutations, i.e., c.731–2A > C and c.786_787insTT (p. (Lys263Leufs*19), which were interpreted as pathogenic according to the ACMG criteria (Zhu et al., 2019). Our results

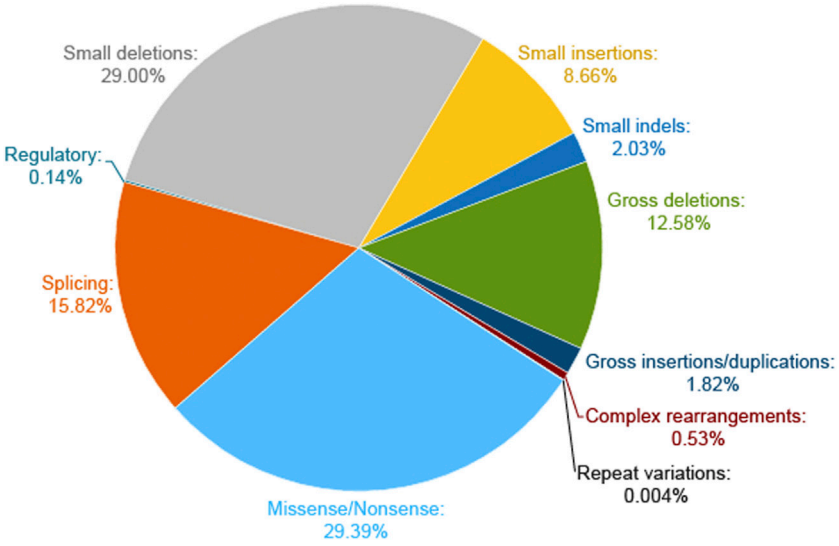


FIGURE 8
Pie chart of *NF1* mutation types on the HGMD website. The micro lesions are made up of missense/nonsense, splicing, regulatory, small deletions, small insertions, and small indels. The gross deletions consist of gross deletions, gross insertions/duplications, and complex rearrangements of repeat variations.

and previous findings have confirmed that the NF1 loss-of-function variant is a major factor leading to NF1 CPT. In the present study, we did not detect the lesion tissue. Zheng et al. (2022) collected the periosteum tissue from the pseudarthrosis site of six patients with NF1-CPT (i.e., probands) and their unaffected parents. They found that five of six NF1-CPT patients (83.3%) had NF1 inactivation in tissues. In addition, we had to admit that WES may not detect all the NF1 variants. For instance, non-coding variants from the regulating area of NF1 may be among the undetected genetic regions (Zheng et al., 2022). Therefore, we recommended using comprehensive detection and analysis of other variants using both the lesion tissue and the blood of patients with and without NF1 CPT.

The Glu^{291} is located within the N-terminal domain of neurofibromin. This truncated variant resulted in the loss of a series of domains including CSRD, GRD, SEC14-PH, and CTD. The functions of these domains might explain the clinical phenotypes in this pedigree to some extent. Although the research on the *NF1* gene has been well studied in recent years, current clinical therapies still mainly rely on symptomatic treatment and surgery, which alleviated clinical symptoms rather than curing the disease. As a monogenic disease, gene therapy is considered a promising therapy. Multiple studies have transduced the *NF1-GRD* gene into different cell lines *via* various vectors and successfully reversed the cell phenotype (Hiatt et al., 2001; Thomas et al., 2006; Bodempudi et al., 2009; Bai et al., 2019).

Given the fact that CPT is a progressive disorder, early diagnosis and intervention are warranted to improve the prognosis of the disease. Owing to the complexity of clinical phenotypes, an extended genetic characterization of this multiple system disorder will be helpful in clinical diagnosis as well as treatment. Precise genetic counseling and gene diagnosis are recommended for CPT patients. If the proband in our study could get early gene detection and be provided with the plaster or cast support and specific nutrition support at an early age, the progression of bone deformities might be slowed down and his prognosis might be better. Therefore, based on the data in this study, we call for attention from pediatric orthopedists, other health care professionals, and patients' caregivers and highly recommended screening for pathogenetic mutations.

Data availability statement

The datasets for this article are not publicly available due to concerns regarding participant/patient anonymity. Requests to access the datasets should be directed to the corresponding authors.

Ethics statement

The studies involving human participants were reviewed and approved by It has been approved by the Local Research Ethics

Committee of the Children's Hospital, Zhejiang University School of Medicine (2022-IRB-148). Written informed consent to participatein this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

Author contributions

Conception and design of the study: JX,YZ, GB, and LH; Data collection: JX, KZ, JL, XH, and YG; Analysis and interpretation of results: JX, YZ, XJ, GB, and LH; Drafting the manuscript: JX, YZ, GB, and LH. All authors reviewed the results and approved the final version of the manuscript.

Funding

This work was supported by the Natural Science Foundation of Zhejiang Province (No. LQ22C070004) and the Basic Public Welfare Research Project of Zhejiang Province of China (LY20H060001).

Acknowledgments

Thanks to the patient and his family for supplying the clinical data and all peer reviewers for their suggestions. We are extremely grateful to all members of the Hu Lab and Bai Lab, past and present, for the interesting discussions and great contributions to the project.

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.

References

Andersen, K. S. (1971). Occurrence of congenital tibial pseudoarthrosis in Denmark 1940-1965. *Nord. Med.* 86, 1395.

Bai, R. Y., Esposito, D., Tam, A. J., Mccormick, F., Riggins, G. J., Wade Clapp, D., et al. (2019). Feasibility of using NF1-GRD and AAV for gene replacement therapy in NF1-associated tumors. *Gene Ther.* 26, 277–286. doi:10.1038/s41434-019-0080-9

Basu, T. N., Gutmann, D. H., Fletcher, J. A., Glover, T. W., Collins, F. S., and Downward, J. (1992). Aberrant regulation of ras proteins in malignant tumour cells from type 1 neurofibromatosis patients. *Nature* 356, 713–715. doi:10.1038/356713a0

Bodempudi, V., Yamoutpoor, F., Pan, W., Dudek, A. Z., Esfandyari, T., Piedra, M., et al. (2009). Ral overactivation in malignant peripheral nerve sheath tumors. *Mol. Cell. Biol.* 29, 3964–3974. doi:10.1128/MCB.01153-08

Brekelmans, C., Hollants, S., de Groote, C., Sohier, N., Maréchal, M., Geris, L., et al. (2019). Neurofibromatosis type 1-related pseudarthrosis: Beyond the pseudarthrosis site. *Hum. Mutat.* 40, 1760–1767. doi:10.1002/humu.23783

Castle, B., Baser, M. E., Huson, S. M., Cooper, D. N., and Upadhyaya, M. (2003). Evaluation of genotype-phenotype correlations in neurofibromatosis type 1. *J. Med. Genet.* 40, e109. doi:10.1136/jmg.40.10.e109

Cichowski, K., and Jacks, T. (2001). NF1 tumor suppressor gene function: Narrowing the GAP. *Cell.* 104, 593–604. doi:10.1016/s0092-8674(01)00245-8

Crawford, A. H. (1986). Neurofibromatosis in children. *Acta Orthop. Scand. Suppl.* 218, 1–60.

Crawford, A. H., and Schorry, E. K. (1999). Neurofibromatosis in children: The role of the orthopaedist. *J. Am. Acad. Orthop. Surg.* 7, 217–230. doi:10.5435/00124635-199907000-00002

de Luca, A., Schirinzi, A., Buccino, A., Bottillo, I., Sinibaldi, L., Torrente, I., et al. (2004). Novel and recurrent mutations in the NF1 gene in Italian patients with neurofibromatosis type 1. *Hum. Mutat.* 23, 629. doi:10.1002/humu.9245

Feng, L., Yunoue, S., Tokuo, H., Ozawa, T., Zhang, D., Patrakitkomjorn, S., et al. (2004). PKA phosphorylation and 14-3-3 interaction regulate the function of neurofibromatosis type I tumor suppressor, neurofibromin. *FEBS Lett.* 557, 275–282. doi:10.1016/s0014-5793(03)01507-2

Granchi, D., Devescovi, V., Baglio, S. R., Magnani, M., Donzelli, O., and Baldini, N. (2012). A regenerative approach for bone repair in congenital pseudarthrosis of the tibia associated or not associated with type 1 neurofibromatosis: Correlation between laboratory findings and clinical outcome. *Cytotherapy* 14, 306–314. doi:10.3109/14653249.2011.627916

Hefti, F., Bollini, G., Dungl, P., Fixsen, J., Grill, F., Ippolito, E., et al. (2000). Congenital pseudarthrosis of the tibia: History, etiology, classification, and epidemiologic data. *J. Pediatr. Orthop. B* 9, 11–15. doi:10.1097/01202412-200001000-00003

Heikkinen, E. S., Poyhonen, M. H., Kinnunen, P. K., and Seppänen, U. I. (1999). Congenital pseudarthrosis of the tibia. Treatment and outcome at skeletal maturity in 10 children. *Acta Orthop. Scand.* 70, 275–282. doi:10.3109/17453679908997807

Hiatt, K. K., Ingram, D. A., Zhang, Y., Bollag, G., and Clapp, D. W. (2001). Neurofibromin GTPase-activating protein-related domains restore normal growth in Nf1-/- cells. *J. Biol. Chem.* 276, 7240–7245. doi:10.1074/jbc.M009202200

Horn, J., Steen, H., and Terjesen, T. (2013). Epidemiology and treatment outcome of congenital pseudarthrosis of the tibia. *J. Child. Orthop.* 7, 157–166. doi:10.1007/s11832-012-0477-0

Izawa, I., Tamaki, N., and Saya, H. (1996). Phosphorylation of neurofibromatosis type 1 gene product (neurofibromin) by cAMP-dependent protein kinase. *FEBS Lett.* 382, 53–59. doi:10.1016/0014-5793(96)00137-8

Jett, K., and Friedman, J. M. (2010). Clinical and genetic aspects of neurofibromatosis 1. *Genet. Med.* 12, 1–11. doi:10.1097/GIM.0b013e3181bf15e3

Kalia, S. S., Adelman, K., Bale, S. J., Chung, W. K., Eng, C., Evans, J. P., et al. (2017). Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): A policy statement of the American College of medical genetics and genomics. *Genet. Med.* 19, 249–255. doi:10.1038/gim.2016.190

Kesireddy, N., Kheireldin, R. K., Lu, A., Cooper, J., Liu, J., and Ebraheim, N. A. (2018). Current treatment of congenital pseudarthrosis of the tibia: A systematic review and meta-analysis. *J. Pediatr. Orthop. B* 27, 541–550. doi:10.1097/BPB.0000000000000524

Kuorilehto, T., Ekholm, E., Nissinen, M., Hietaniemi, K., Hiltunen, A., Paavolainen, P., et al. (2006). NF1 gene expression in mouse fracture healing

and in experimental rat pseudarthrosis. *J. Histochem. Cytochem.* 54, 363–370. doi:10.1369/jhc.5A6784.2005

Mangoura, D., Sun, Y., Li, C., Singh, D., Gutmann, D. H., Flores, A., et al. (2006). Phosphorylation of neurofibromin by PKC is a possible molecular switch in EGF receptor signaling in neural cells. *Oncogene* 25, 735–745. doi:10.1038/sj.onc.1209113

Marchuk, D. A., Saulino, A. M., Tavakkol, R., Swaroop, M., Wallace, M. R., Andersen, L. B., et al. (1991). cDNA cloning of the type 1 neurofibromatosis gene: complete sequence of the NF1 gene product. *Genomics* 11, 931–940. doi:10.1016/0888-7543(91)90017-9

Mladenov, K. V., Spiro, A. S., Krajewski, K. L., Stücker, R., and Kunkel, P. (2020). Management of spinal deformities and tibial pseudarthrosis in children with neurofibromatosis type 1 (NF-1). *Childs Nerv. Syst.* 36, 2409–2425. doi:10.1007/s00381-020-04775-4

Mousley, C. J., Tyeryar, K. R., Vincent-Pope, P., and Bankaitis, V. A. (2007). The Sec14-superfamily and the regulatory interface between phospholipid metabolism and membrane trafficking. *Biochim. Biophys. Acta* 1771, 727–736. doi:10.1016/j.bbalip.2007.04.002

Philpott, C., Tovell, H., Frayling, I. M., Cooper, D. N., and Upadhyaya, M. (2017). The NF1 somatic mutational landscape in sporadic human cancers. *Hum. Genomics* 11, 13. doi:10.1186/s40246-017-0109-3

Sabbagh, A., Pasmant, E., Imbard, A., Luscan, A., Soares, M., Blanché, H., et al. (2013). NF1 molecular characterization and neurofibromatosis type I genotype-phenotype correlation: The French experience. *Hum. Mutat.* 34, 1510–1518. doi:10.1002/humu.22392

Sharma, R., Wu, X., Rhodes, S. D., Chen, S., He, Y., Yuan, J., et al. (2013). Hyperactive Ras/MAPK signaling is critical for tibial nonunion fracture in neurofibromin-deficient mice. *Hum. Mol. Genet.* 22, 4818–4828. doi:10.1093/hmg/ddt333

Thomas, S. L., Deadwyler, G. D., Tang, J., Stubbs, E. B., JR., Muir, D., Hiatt, K. K., et al. (2006). Reconstitution of the NF1 GAP-related domain in NF1-deficient human Schwann cells. *Biochem. Biophys. Res. Commun.* 348, 971–980. doi:10.1016/j.bbrc.2006.07.159

Tokuo, H., Yunoue, S., Feng, L., Kimoto, M., Tsuji, H., Ono, T., et al. (2001). Phosphorylation of neurofibromin by cAMP-dependent protein kinase is regulated via a cellular association of N(G), N(G)-dimethylarginine dimethylaminohydrolase. *FEBS Lett.* 494, 48–53. doi:10.1016/s0014-5793(01)02309-2

Trovó-Marqui, A. B., and Tajara, E. H. (2006). Neurofibromin: A general outlook. *Clin. Genet.* 70, 1–13. doi:10.1111/j.1399-0004.2006.00639.x

van Royen, K., Brems, H., Legius, E., Lammens, J., and Laumen, A. (2016). Prevalence of neurofibromatosis type 1 in congenital pseudarthrosis of the tibia. *Eur. J. Pediatr.* 175, 1193–1198. doi:10.1007/s00431-016-2757-z

Vander Have, K. L., Hensinger, R. N., Caird, M., Johnston, C., and Farley, F. A. (2008). Congenital pseudarthrosis of the tibia. *J. Am. Acad. Orthop. Surg.* 16, 228–236. doi:10.5435/00124635-200804000-00006

Viskochil, D., Buchberg, A. M., Xu, G., Cawthon, R. M., Stevens, J., Wolff, R. K., et al. (1990). Deletions and a translocation interrupt a cloned gene at the neurofibromatosis type 1 locus. *Cell.* 62, 187–192. doi:10.1016/0092-8674(90)90252-a

Viskochil, D. (2002). Genetics of neurofibromatosis 1 and the NF1 gene. *J. Child. Neurol.* 17, 562–570. ; discussion 571-2, 646-51. doi:10.1177/088307380201700804

World Medical Association (2013). World medical association declaration of helsinki: Ethical principles for medical research involving human subjects. *Jama* 310, 2191–2194.

Xu, G. F., O'Connell, P., Viskochil, D., Cawthon, R., Robertson, M., Culver, M., et al. (1990). The neurofibromatosis type 1 gene encodes a protein related to GAP. *Cell.* 62, 599–608. doi:10.1016/0092-8674(90)90024-9

Young, H., Hyman, S., and North, K. (2002). Neurofibromatosis 1: Clinical review and exceptions to the rules. *J. Child. Neurol.* 17, 613–621. ; discussion 627-9, 646-51. doi:10.1177/088307380201700812

Zheng, Y., Zhu, G., Liu, Y., Zhao, W., Yang, Y., Luo, Z., et al. (2022). Case series of congenital pseudarthrosis of the tibia unfulfilling neurofibromatosis type 1 diagnosis: 21% with somatic NF1 haploinsufficiency in the periosteum. *Hum. Genet.* 141, 1371–1383. doi:10.1007/s00439-021-02429-2

Zhu, G., Zheng, Y., Liu, Y., Yan, A., Hu, Z., Yang, Y., et al. (2019). Identification and characterization of NF1 and non-NF1 congenital pseudarthrosis of the tibia based on germline NF1 variants: Genetic and clinical analysis of 75 patients. *Orphanet J. Rare Dis.* 14, 221. doi:10.1186/s13023-019-1196-0

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 17 October 2022
DOI 10.3389/fgene.2022.948505

Check for updates

OPEN ACCESS

EDITED BY
Xianyong Yin,
Department of Biostatistics, University of Michigan, United States

REVIEWED BY
Xu-jie Zhou,
First Hospital, Peking University, China
Ashish Kapoor,
University of Texas Health Science Center at Houston, United States

*CORRESPONDENCE
Gaafar Ragab,
GaafarR@gmail.com
Swapan K. Nath,
Swapan-Nath@omrf.org

[§]These authors co-directed the project

[†]These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

RECEIVED 19 May 2022
ACCEPTED 30 September 2022
PUBLISHED 17 October 2022

CITATION
Elghzaly AA, Sun C, Looger LL, Hirose M, Salama M, Khalil NM, Behiry ME, Hegazy MT, Hussein MA, Salem MN, Eltoraby E, Tawhid Z, Alwasefy M, Allam W, El-Shiekh I, Elserafy M, Abdelnaser A, Hashish S, Shebl N, Shahba AA, Elgirby A, Hassab A, Refay K, El-Touchy HM, Youssef A, Shabacy F, Hashim AA, Abdelzaher A, Alshebini E, Fayez D, El-Bakry SA, Elzohri MH, Abdelsalam EN, El-Khamisy SF, Ibrahim S, Ragab G and Nath SK (2022), Genome-wide association study for systemic lupus erythematosus in an egyptian population.
Front. Genet. 13:948505.
doi: 10.3389/fgene.2022.948505

COPYRIGHT
© 2022 Elghzaly, Sun, Looger, Hirose, Salama, Khalil, Behiry, Hegazy, Hussein, Salem, Eltoraby, Tawhid, Alwasefy, Allam, El-Shiekh, Elserafy, Abdelnaser, Hashish, Shebl, Shahba, Elgirby, Hassab, Refay, El-Touchy, Youssef, Shabacy, Hashim, Abdelzaher, Alshebini, Fayez, El-Bakry, Elzohri, Abdelsalam, El-Khamisy, Ibrahim, Ragab and Nath. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Genome-wide association study for systemic lupus erythematosus in an egyptian population

Ashraf A. Elghzaly[1†], Celi Sun[2†], Loren L. Looger[3†], Misa Hirose[4], Mohamed Salama[5], Noha M. Khalil[6], Mervat Essam Behiry[6], Mohamed Tharwat Hegazy[6], Mohamed Ahmed Hussein[6], Mohamad Nabil Salem[7], Ehab Eltoraby[8], Ziyad Tawhid[1], Mona Alwasefy[1], Walaa Allam[9], Iman El-Shiekh[9], Menattallah Elserafy[9], Anwar Abdelnaser[5], Sara Hashish[5], Nourhan Shebl[5], Abeer Abdelmonem Shahba[10], Amira Elgirby[11], Amina Hassab[12], Khalida Refay[13], Hanan Mohamed El-Touchy[13], Ali Youssef[14], Fatma Shabacy[14], Abdelkader Ahmed Hashim[15], Asmaa Abdelzaher[16], Emad Alshebini[17], Dalia Fayez[18], Samah A. El-Bakry[18], Mona H. Elzohri[19], Eman Nagiub Abdelsalam[20], Sherif F. El-Khamisy[21,22,9], Saleh Ibrahim[4], Gaafar Ragab[6*§] and Swapan K. Nath[2*§]

[1]Department of Clinical Pathology, Faculty of Medicine, Mansoura University, El-Mansoura, Egypt, [2]Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK, United States, [3]Department of Neurosciences, Howard Hughes Medical Institute, University of California, San Diego, San Diego, CA, United States, [4]Division of Genetics, Lübeck Institute of Experimental Dermatology, University of Lübeck, Lübeck, Germany, [5]Institute of Global Health and Human Ecology, The American University in Cairo, New Cairo, Egypt, [6]Rheumatology and Clinical Immunology Unit, Department of Internal Medicine, Faculty of Medicine, Cairo University, Cairo, Egypt, [7]Department of Internal Medicine, Faculty of Medicine, Beni-Suef University, Beni Suef, Egypt, [8]Department of Internal Medicine, Faculty of Medicine, Mansoura University, El-Mansoura, Egypt, [9]Center for Genomics, Helmy Institute for Medical Sciences, Zewail City of Science and Technology, Giza, Egypt, [10]Department of Internal Medicine, Faculty of Medicine, Tanta University, Tanta, Egypt, [11]Department of Internal Medicine, Faculty of Medicine, Alexandria University, Bab Sharqi, Egypt, [12]Department of Clinical Pathology, Faculty of Medicine, Alexandria University, Bab Sharqi, Egypt, [13]Department of Internal Medicine, Faculty of Medicine, Al-Azhar University, Cairo, Egypt, [14]Department of Rheumatology and Immunology, Faculty of Medicine, Benha University Hospital, Benha, Egypt, [15]Department of Internal Medicine, Faculty of Medicine, South Valley University, Qena, Egypt, [16]Department of Clinical Pathology, Faculty of Medicine, South Valley University, Qena, Egypt, [17]Department of Internal Medicine, Faculty of Medicine, Menoufia University, Al Minufiyah, Egypt, [18]Rheumatology and Clinical Immunology Unit, Department of Internal Medicine, Faculty of Medicine, Ain Shams University, Cairo, Egypt, [19]Department of Internal Medicine, Faculty of Medicine, Assiut University, Asyut, Egypt, [20]Department of Clinical Pathology, Faculty of Medicine, Assiut University, Asyut, Egypt, [21]The Healthy Lifespan Institute, University of Sheffield, Sheffield, United Kingdom, [22]The Institute of Cancer Therapeutics, University of Bradford, Bradford, United Kingdom

Systemic lupus erythematosus (SLE) susceptibility has a strong genetic component. Genome-wide association studies (GWAS) across trans-ancestral populations show both common and distinct genetic variants of susceptibility across European and Asian ancestries, while many other ethnic populations remain underexplored. We conducted the first SLE GWAS on Egyptians—an admixed North African/Middle Eastern population—using

537 patients and 883 controls. To identify novel susceptibility loci and replicate previously known loci, we performed imputation-based association analysis with 6,382,276 SNPs while accounting for individual admixture. We validated the association analysis using adaptive permutation tests ($n = 10^9$). We identified a novel genome-wide significant locus near *IRS1/miR-5702* ($P_{corrected} = 1.98 \times 10^{-8}$) and eight novel suggestive loci ($P_{corrected} < 1.0 \times 10^{-5}$). We also replicated ($P_{perm} < 0.01$) 97 previously known loci with at least one associated nearby SNP, with *ITGAM, DEF6-PPARD* and *IRF5* the top three replicated loci. SNPs correlated ($r^2 > 0.8$) with lead SNPs from four suggestive loci (*ARMC9, DIAPH3*, *IFLDT1,* and *ENTPD3*) were associated with differential gene expression ($3.5 \times 10^{-95} < p < 1.0 \times 10^{-2}$) across diverse tissues. These loci are involved in cellular proliferation and invasion—pathways prominent in lupus and nephritis. Our study highlights the utility of GWAS in an admixed Egyptian population for delineating new genetic associations and for understanding SLE pathogenesis.

KEYWORDS

GWAS, lupus, admixture, Egypt, imputation

1 Introduction

Systemic lupus erythematosus (SLE) is a chronic, complex, multi-system autoimmune disease with substantial mortality and morbidity. Prevalence, severity, and sub-clinical manifestations vary significantly across ethnically diverse populations—with an outsized burden on individuals with African, Hispanic, and Asian backgrounds compared to Caucasians (Carter et al., 2016; Lewis and Jawad, 2017; Pisetsky et al., 2017). Additionally, SLE shows remarkable gender bias, affecting ~9–10 times as many women as men.

In addition to epigenetic and environmental contributions, SLE has a very strong genetic component. This is evidenced by high heritability (~66%), familial clustering (sibling recurrence risk ratio ~30), twin studies (~10 times higher concordance rate in monozygotic *versus* dizygotic twins) (Deapen et al., 1992; Block, 2006; Mak and Tay, 2014; Kuo et al., 2015), and by a growing number of susceptibility loci identified by genome-wide association studies (GWAS) and high-density candidate gene studies [e.g., ImmunoChip (Cortes and Brown, 2011)]. Despite the identification of ~180 SLE susceptibility loci (Bentham et al., 2015; Sun et al., 2016; Langefeld et al., 2017; Molineros et al., 2017; Julia et al., 2018; Wang et al., 2018; Tangtanatakul et al., 2020; Ha et al., 2022), these account for only ~30% of genetic heritability (Yin et al., 2020), indicating that many genes and pathways remain incompletely mapped or even undiscovered.

The statistical power and locus resolution of GWAS and candidate-gene studies can be increased in several ways: larger sample size, deeper genotyping of existing samples, improved imputation and statistical fine-mapping, and most importantly, introducing samples with different allele and haplotype usage than existing samples. Critically, most lupus studies—and most genetic studies in general—are built mainly upon individuals with European (and to a lesser extent Asian) ancestries (Ha et al., 2022). However, many alleles occur at low frequencies in these ethnicities, preventing statistically meaningful measurement of their effects. The use of underrepresented ethnicities with different allelic usage and recombination hotspots is the single most effective way to increase association power (Wojcik et al., 2019). In addition to providing sufficient power to resolve additional ancestry-independent signals, such studies also illuminate ancestry-specific signals (Goulielmos et al., 2018)—wherein a locus is risk only in a specific ethno-genetic background. Such studies are particularly important in autoimmune diseases like lupus, which afflict non-white ethnicities at much higher rates than whites.

Lack of diversity in the available human genomes deposited in public databases limits our understanding of the genetic underpinnings of complex traits, hinders precision medicine, and contributes to health disparities (Popejoy and Fullerton, 2016; Ben-Eghan et al., 2020). Africans and Middle Easterners are among the populations the most underrepresented in genetic association studies, despite comprising over 1.5 billion people and coming from the birthplace of humanity. Moreover, Africa is home to more genetic diversity than the rest of the world combined (Tishkoff et al., 2009) and as such constitutes a woefully underutilized pool for genetic discoveries. Meanwhile, the Middle East and North Africa were both the cradle of civilization and the largest crossroads of migration in the ancient world, bringing together Africa, Europe, and Asia as empires rose and fell. A recent population admixture study (Schuenemann et al., 2017) showed that Ancient Egyptians were more closely related to Middle Easterners and Europeans than to Africans, with African ancestry increasing after the fall of the Roman Empire. In SLE, striking racial/ethnic differences exist in incidence, disease course and clinical manifestations; genetic

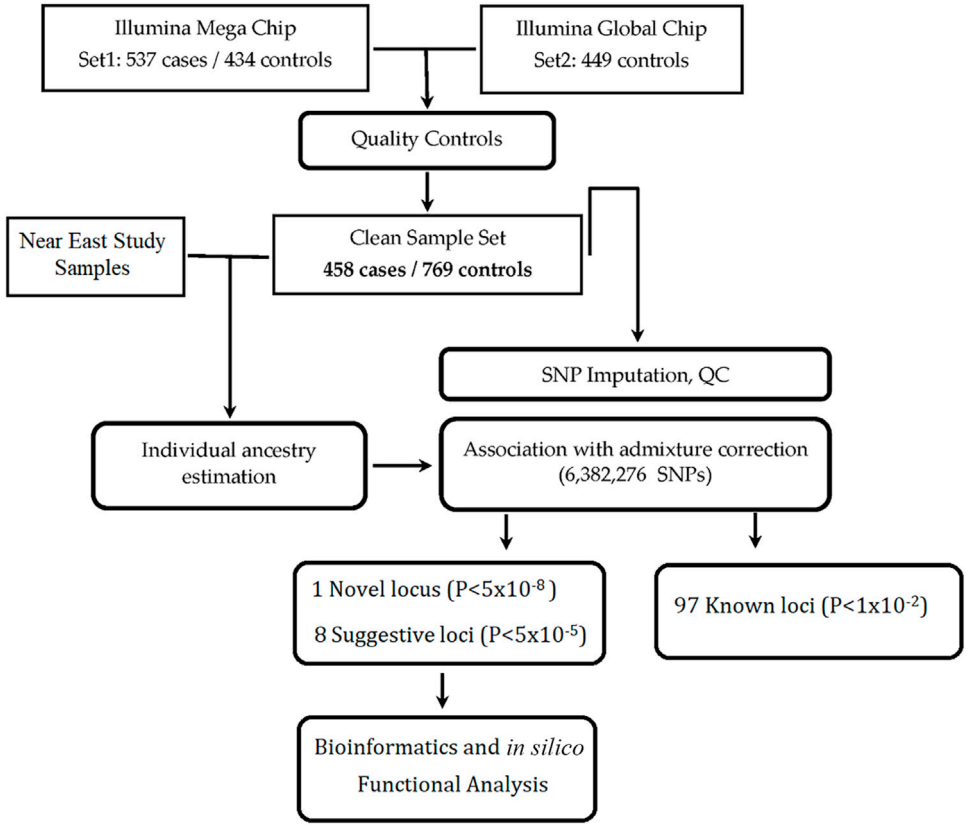


FIGURE 1
Overview and study design.

studies will be helpful to delineate these disparities. However, very few genetic association studies on Egyptians have been performed thus far—and most feature only a few SNPs from a handful of candidate genes (Elghzaly et al., 2015). Therefore, a large-scale genome-wide wide association study is required to understand the SLE genetic landscape in Egyptians—here, we perform the first such study. The major findings of our study are: 1) nine newly uncovered SLE susceptibility loci with one genome-wide significant, 2) replication of several previously known SLE loci, and 3) remarkable similarities in minor allele frequencies at risk loci between Egyptians and Europeans.

2 Materials and methods

2.1 Systemic lupus erythematosus patients and controls

The overview and design of the study is shown in Figure 1. Patients were informed of the nature of the study, and only those who gave their consent were included in the study. SLE was diagnosed according to the 1997 update of the Revised American College of Rheumatology (ACR) classification criteria for SLE (Hochberg, 1997) or the "Systemic Lupus International Collaborating Clinic" (SLICC) Criteria for Classification of Systemic Lupus Erythematosus (Petri et al., 2012). Our rheumatologists used both diagnostic criteria, because we had patients of long disease history with established diagnosis based on earlier criteria that preceded the SLICC standard. Patients who were identified as SLE positive after the introduction of the SLICC criteria were diagnosed accordingly and the two subsets, being included in each center's registries, were recruited for the study as definite cases of SLE. Note that we lack ACR/SLICC information and some demographic data for 26 SLE patients who have left physicians' care and are no longer available for verification. Patients' recruitment, clinical examinations, diagnostic laboratory investigations, and data collection were all performed in their corresponding medical centers by expert rheumatology teams and in specialized university laboratories. Data was collected in an *ad hoc* unified Excel spreadsheet and reviewed by senior rheumatology experts. Exclusion criteria included the presence of overlap features or the presence of

other autoimmune diseases except for Sjögren's syndrome. The control group was selected from the same geographical location as the corresponding center. All participants provided informed consent, and the study was approved by the Institutional Review Boards from Oklahoma Medical Research Foundation (OMRF), Oklahoma City, United States (IRB approval number 14–16), and Mansoura University Ethics committee/IRB (MFM-IRB code number R/16.04.81). The recruiting hospitals/universities included Cairo University, Mansoura University, Tanta University, Alazhar University–females only, Beni-Suef University, Benha University, Menofia University, Alexandria University, Ain Shams University, South Valley University, and Assuit University.

Initially, 537 Egyptian SLE patients and 434 controls were recruited for the GWAS study, and DNA samples were extracted from 5 ml whole blood samples collected in EDTA-containing tubes and stored at −20°C. DNA extraction for all samples was done in Germany, according to manufacturer's instructions using the Qiamp DNA Blood Mini Kit (QIAGEN, Hilden, Germany). To increase statistical power, 449 additional unaffected Egyptian controls [already genotyped in Germany for another GWAS study (Bejaoui et al., 2019)] were added to the study. These controls were recruited from the blood banks of Cairo and Mansoura Universities and approved by the IRBs of both Universities (as mentioned above).

2.2 Demographics

Gender and age distributions for SLE cases and controls (after quality control, described below) are shown in Supplementary Table S1. Overall, females constituted 90% and 64% of cases and controls, respectively. Average ages of cases and controls are comparable. The distribution of major ACR criteria within SLE patients are shown in Supplementary Table S1. Almost all (99%) SLE patients were positive for anti-nuclear antibodies (ANA).

2.3 Genotyping and quality control

Initial samples (537 SLE patients and 434 controls) were genotyped on the Illumina Multi-Ethnic Genotyping Array (MEGA) at the genotyping core facility of OMRF, Oklahoma City, United States. The 449 additional out-of-study controls were already genotyped on the Illumina Infinium Global Screening Array. The out-of-study controls have no genotyping data on the X-chromosome; thus we decided to focus only on autosomal loci. We hope to obtain this missing genotyping data for use in future studies on X-linked traits. We performed strict quality control (QC) on genotyped SNPs in each dataset separately (cases and two sets of controls) as follows: (a) we excluded SNPs with bad clusters due to poor genotyping calls, (b) we removed SNPs with missing genotype rate ≥ 0.05, and (c) we filtered out SNPs out of Hardy-Weinberg equilibrium ($p < 0.0001$) and/or minor allele frequency (MAF) < 0.5%. After QC, 295,981 autosomal (chromosomes 1–22) SNPs that overlapped both genotyping arrays remained for further analysis. Of the 537 cases and 883 controls, the GCTA algorithm (Yang et al., 2011) selected 500 cases and 815 controls as being unrelated from one another; the others were excluded. As a final QC step, 42 cases and 46 controls were removed from the cohort due to being outliers from principal components analysis—leaving 458 cases and 769 controls for this study. The inflation factor λ was calculated from the case-control association test on the final, clean cohort (458 cases and 769 controls).

2.4 Statistical power for detecting association

To estimate the ability of our GWAS cohort to support statistically significant associations, we performed a statistical power calculation. The Genetic Association Study Power Calculator server (Johnson and Abecasis, 2017), derived from the CaTS power calculator for two-stage association studies (Skol et al., 2006), was used. We used our final cohort (458 cases, 769 controls) to assess statistical power, while setting the parameter prevalence for SLE to 0.3% (Wang Y. F. et al., 2021).

2.5 Imputation for GWAS samples

We used the Michigan Imputation Server (Das et al., 2016) for imputation on the 22 autosomes. 1,000 Genomes Project data (Phase 3 Integrated Release Version 5 Haplotypes) was used as the reference panel. After imputation, we performed strict QC on post-imputed SNPs as described above (Hardy-Weinberg equilibrium $p > 0.0001$ in controls and/or MAF ≥ 0.5%). About 6.36 million post-imputed SNPs with high imputation quality (Rsq > 0.7 for MAF ≥ 3%) were used for downstream analysis.

2.6 Admixture-corrected association analysis with imputed data

To quantify individual ancestry proportion in our Egyptian cohort, we complemented our dataset with published samples from the ancient Near East (Lazaridis et al., 2016) (the historical Fertile Crescent and Levant regions, corresponding to modern-day Egypt, Turkey, Iran, and surrounding nations.) The Near East samples include 294 ancient and 2,068 modern individuals spanning a range of nationalities and ethnicities. In both the Egyptian dataset and Near East public dataset, we excluded SNPs with A/T or C/G alleles and identified ~33,000 common

unrelated SNPs (MAF > 5% and linkage disequilibrium, LD r^2 < 0.2) for analysis. Principal components analysis was performed with GCTA (Yang et al., 2011), and genetic ancestry composition was estimated with ADMIXTURE (Alexander et al., 2009).

We performed imputation-based association analyses using mach2dat (Li et al., 2009) before and after admixture correction. (A/T and C/G alleles were recovered during imputation.) ADMIXTURE uses a maximum-likelihood approach to determine admixture proportions of individuals by assuming there are K hypothetical underlying populations among them. We selected $K = 5$ as the best model for the number of ancestral populations and took the first four ancestry proportions as covariates for admixture correction. The inflation factor (λ) was estimated using all admixture-corrected association results from QC-passed SNPs.

We performed multiple-testing correction using permutation (Gao, 2011)—considered the gold standard in GWAS studies. Note that we specifically avoided Bonferroni correction as this assumes independence, which is failed when any SNPs are in linkage disequilibrium—thus producing unacceptably conservative associations (Johnson et al., 2010). Besides, Bonferroni correction suffers several shortcomings, most notably low statistical power for rejecting the null hypothesis (Cohen, 1990, 1994; Perneger, 1998; Nakagawa, 2004) and also performing large amounts of unnecessary testing (Fadista et al., 2016). Critically, permutation analysis controls for low sample-size effects, as all permutations have identical n as the proposal—comparisons are between permutations with identical statistical power *a priori* to produce an association. Thus, we do not fear spurious associations despite our relatively small sample sizes.

2.7 Replication of previously known variants and loci

To replicate previously reported SLE susceptibility loci, we followed a recent study (Ha et al., 2022) that collated 179 statistically-independent non-human leukocyte antigen (HLA) loci as underlying SLE susceptibility. At each risk locus, we defined a boundary as ± 150 kb around the nominated SNP and collated the full set of SNPs in these intervals. Then, we compared this (enlarged) SNP set with the admixture-corrected association results. Using an ancestry-corrected permutation-based analysis, we set the threshold for replication at $P_{perm} < 0.01$—stricter than the more commonly used 0.05 threshold, to better exclude false positive associations.

2.8 SNP functional annotation

Functional interpretation of the associated SNP under any GWAS peak is of critical importance in follow-up analysis. Like most other GWAS studies, many of our identified SNPs concentrate in non-coding DNA, e.g., transcriptional regulatory regions (promoters, enhancers) and non-coding RNA. Therefore, analyzing tissue-specific effects of expression quantitative trait loci (eQTLs) is a promising approach. We used Qtlizer (Munz et al., 2020) for the top SNPs in the novel loci to obtain their eQTLs from several public databases, including GTEx v7 (Consortium, 2013), GEUVADIS (Lappalainen et al., 2013), GRASP (Leslie et al., 2014), Haploreg (Ward and Kellis, 2012), SCAN (Gamazon et al., 2009), seeQTL (Xia et al., 2011), Blood eQTL Browser (Westra et al., 2013), pGWAS (Suhre et al., 2017), ExSNP (Yu et al., 2016), and BRAINEAC (Ramasamy et al., 2014). Qtlizer allowed exploration of QTL data for a given list of variants (indels and SNPs) in a fast and efficient manner by integrating many QTL datasets.

2.9 Annotation of long non-coding RNA targets and microRNAs

Target sites of lncRNAs were predicted using the LncRRIsearch server (http://rtools.cbrc.jp/LncRRIsearch) (Fukunaga et al., 2019), built upon RIblast (Fukunaga and Hamada, 2017), which computationally identifies lncRNA-mRNA interactions through a seed-and-extension approach. For miRNAs, we identified similar sequences in the human genome (GRCh37/hg19) with BLAT on the UCSC Genome Browser (GRCh37/hg19).

3 Results

3.1 Study population and statistical power

After rigorous QC, 458 cases and 769 controls were identified by GCTA (Yang et al., 2011) as unrelated samples and were used for association (Figure 1, Section 2). Due to the small sample size, our study had moderate power (68%) for detection of significant ($p < 1 \times 10^{-5}$) variants with odds ratio (OR) < 1.5 (Supplementary Figure S1). However, for common (minor allele frequency, MAF ≥ 10%) SNPs with large effect (OR ≥ 1.6), our study had high (up to 90%) power. The inflation factor (λ) between cases and controls was 1.06; the QQ plot is shown in Supplementary Figure S2.

3.2 Genome-wide association of SLE

To discover novel SLE susceptibility loci and to replicate previously known SLE risk loci in our Egyptian cohort, we performed a genome-wide imputation-based association study with 6,382,276 SNPs while adjusting for individual admixture (see Section 2). The Manhattan plot showing -$\log_{10}$ p values for SNP associations for our GWAS is shown in Figure 2.

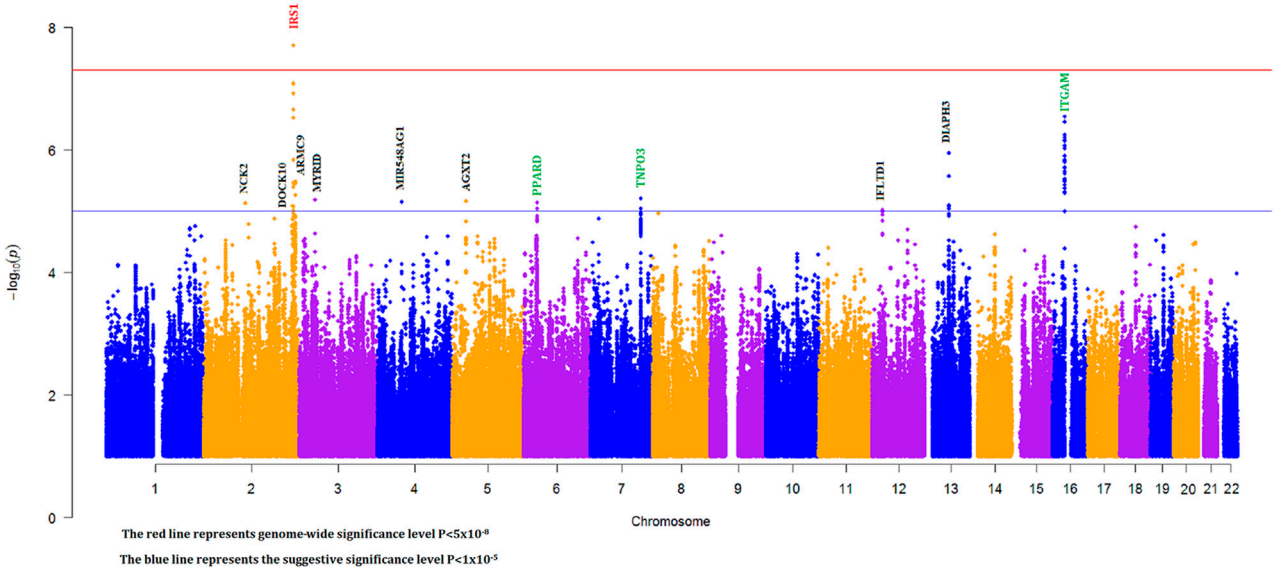


FIGURE 2
Manhattan plot for imputation-based association with individual ancestry correction (green, black, and red gene names represent known, novel suggestive, and novel genome-wide significant loci, respectively).

3.2.1 Discovery of novel SLE loci

We identified 28 SNPs in nine novel loci with $P_{corrected} < 1 \times 10^{-5}$ (Table 1; Supplementary Table S2; Figure 3). The most significant signal, which passed genome-wise significance (rs5839171, $P_{corrected} = 1.98 \times 10^{-8}$, $P_{perm} = 3.30 \times 10^{-8}$), is downstream of the Insulin receptor substrate 1 gene (*IRS1*) and miR-5702, a microRNA involved in cellular proliferation and cancer risk (Li K. et al., 2019). IRS1 is a signaling adapter protein linking activity of insulin and insulin-like growth factor receptors to intracellular signaling cascades, most notably the PI3K/Akt and Erk MAP kinase pathways (Leslie et al., 2014). *IRS1* has been flagged as a risk locus for insulin resistance (Rung et al., 2009) and type II diabetes (Rung et al., 2009; Voight et al., 2010), and also several cancers (Slattery et al., 2004). The best characterized target of miR-5702 is Zinc finger E-box-binding homeobox 1 (*ZEB1*) (Zhang et al., 2018), a transcription factor critically involved in T-lymphocyte differentiation and interleukin-2 signaling (Guan et al., 2018). At this locus, another nearby (1.7 kb) SNP, rs10804331 ($P_{corrected} = 8.5 \times 10^{-8}$), is a significant eQTL for *IRS1* in $CD4^+$ naïve T-cells ($p = 0.002$) and $CD16^+$ neutrophils ($p = 0.01$) (Supplementary Table S3). This SNP is also a significant mQTL for methylated CpG cg07514207 in the *IRS1* promoter (Chen et al., 2016), potentially modulating transcription.

The second signal (rs2265634; $P_{corrected} = 1.14 \times 10^{-6}$) is between Diaphanous homolog 1 (*DIAPH3*), which modulates mTOR signaling and cellular proliferation (Wan et al., 2021), and Tudor domain-containing protein 3 (*TDRD3*), a chromatin modulator also involved in proliferation (Morettin et al., 2017); it is also close to long non-coding RNA *LINC00434*. *LINC00434* appears to target several genes involved in cancer and autoimmunity (Supplementary Table S4). This SNP is a significant ($p = 4.78 \times 10^{-8}$) expression quantitative trait locus (eQTL) for *DIAPH3*, which was a nominally significant ($p = 7.23 \times 10^{-3}$) SLE risk locus in a previous GWAS study (Lee et al., 2014).

The third signal (rs6761645; $P_{corrected} = 3.35 \times 10^{-6}$) is upstream of Armadillo repeat containing 9 [*ARMC9*, encoding KU-MEL-1 (Kiniwa et al., 2001)], a ciliary basal body protein, mutations of which underlie the systemic inflammatory disorder Vogt-Koyanagi-Harada disease (VKH) (Ohno et al., 2019) and the neurodevelopmental disorder Joubert syndrome (Van De Weghe et al., 2017). KU-MEL-1, over-expressed in melanocytes, appears to be a primary autoantigen in VKH (Otani et al., 2006) and is also associated with vitiligo and autoimmune uveitis. These top three hits were all located in active chromatin, with rs6761645, in particular, overlapping promoter signals in 17 tissues and enhancers in 12 tissues (Ward and Kellis, 2012); having 14 significant eQTL hits (Supplementary Table S3); and being associated with transcription factor binding sites for Hypoxia-inducible factor 1 (HIF1), a critical regulator of inflammation and the DNA damage response (Dehne and Brüne, 2009; Ramachandran et al., 2021). rs6761645 is an eQTL for *PSMD1* (26S proteasome non-ATPase regulatory subunit 1)

TABLE 1 Association (genome-wide significant and suggestive) signals at the lead SNPs with $p < 1.0x10^{-5}$.

Chr	SNP	Position (bp, hg19)	Nearest genes	A1/A2	FA/FU	P_uncorrected	P_corrected	#Sig SNPs	P_perm_admix	OR(95%CI)
2	rs5839171	227,405,634	IRS1/MIR5702	G/GA	0.492/0.597	1.70E-07	1.98E-08	12	3.30E-08	0.63(0.53–0.75)
13	rs2265634	60,812,386	LINC00434, DIAPH3	G/A	0.56/0.47	6.66E-06	1.14E-06	5	3.80E-06	1.49(1.25–1.77)
2	rs6761645	232,055,153	ARMC9	A/T	0.183/0.254	4.06E-06	3.35E-06	4	9.00E-07	0.57(0.45–0.73)
3	rs7633684	40,422,493	ENTPD3, ENTPD3-AS1	T/A	0.803/0.875	1.29E-06	6.66E-06	1	5.80E-06	0.56(0.45–0.71)
5	rs148600009	35,023,218	AGXT2	A/ ATTCT	0.201/0.143	4.72E-05	6.88E-06	1	1.31E-05	1.65(1.3–2.1)
4	rs17090343	61,873,612	MIR548AG1	C/T	0.074/0.032	4.33E-06	7.18E-06	1	6.30E-06	2.45(1.67–3.59)
2	rs6743358	106,245,742	NCK2	G/A	0.504/0.411	4.76E-06	7.56E-06	1	6.20E-06	1.48(1.25–1.76)
2	rs11462616	225,900,189	DOCK10	AT/A	0.289/0.377	8.56E-06	8.27E-06	1	1.63E-05	0.67(0.56–0.8)
12	rs12817138	25,826,394	IFLTD1	G/A	0.552/0.638	9.20E-06	9.69E-06	2	7.10E-06	0.67(0.56–0.8)

(Supplementary Table S3), another proliferation marker (Okumura et al., 2017).

The five other novel signals localized to *ENTPD3/ENTPD3-AS1* (a tumor suppressor (Li M. et al., 2019) that strongly signals through HIF1 (Wang J. et al., 2021) and its associated antisense RNA), *NCK2* [a critical immune adaptor protein linking B-cell receptor activation to PI3K signaling (Cannons et al., 2013)], *IFLTD1* [a ciliary organization protein and cancer risk gene (Wang et al., 2005), also known as *LMNTD1* and *PAS1C1*], *AGXT2* [a risk locus of premature myocardial infarct (Yuanfeng et al., 2016)], *DOCK10* [critical in B-cell activation and proliferation (Yelo et al., 2008)], and the microRNA *MIR548AG1*.

Given the prominent role played by microRNAs in cancer (Calin and Croce, 2006) and autoimmunity (Xiao and Rajewsky, 2009), and the fact that miR-548ag1 had the largest effect size of our novel loci [odds ratio = 2.45 (95% confidence interval 1.67–3.59)], we closely examined the apparent involvement of this miRNA in SLE. The miR-548 family was first discovered in colorectal expression arrays (Cummins et al., 2006) and was expanded through bioinformatic searches for miRNAs overlapping repetitive elements (Piriyapongsa and Jordan, 2007). Both the microRNA family itself and putative target sites derive from the Made1 repetitive element; these sequences form very stable hairpin structures of ~37 bp in length. miR-548ag1 was discovered from a B-cell tumor line (Jima et al., 2010), suggesting immune involvement. Many other family members are implicated in cancer (Shi et al., 2015; Ke et al., 2016) and autoimmunity (Yu et al., 2017; Cakmak Genc et al., 2018; Li et al., 2018). To further elucidate the biological roles of miR-548ag1, we searched the human genome for miR-548ag1-like sequences (see Section 2), restricting ourselves to loci exhibiting primate-specific insertions [consistent with other verified miR-548 family members (Piriyapongsa and Jordan, 2007)]. We found closely related sequences in UTRs and introns of ten genes—usually overlapping enhancers and/or transcription factor binding sites (Supplementary Table S5). Strikingly, all ten genes have clear immune involvement (Supplementary Table S5). Five of the ten hits are expressed sequence tags (ESTs) of ~90 bp–100 bp and almost certainly represent previously unannotated members of the miR-548 family. Of the remaining five, it is not immediately clear which are microRNAs themselves and which are potential miRNA target sites. Together, these results emphasize the immune system relevance of our SLE risk locus miR-548ag1 and related miR-548 family members.

3.2.2 Replication of known SLE loci

We compared our association results with 179 SLE known loci (see Section 2) and found 97 known loci replicated with at least one significantly ($P_{perm} < 0.01$) associated SNP. Whenever available, we showed LD with the top published variant to

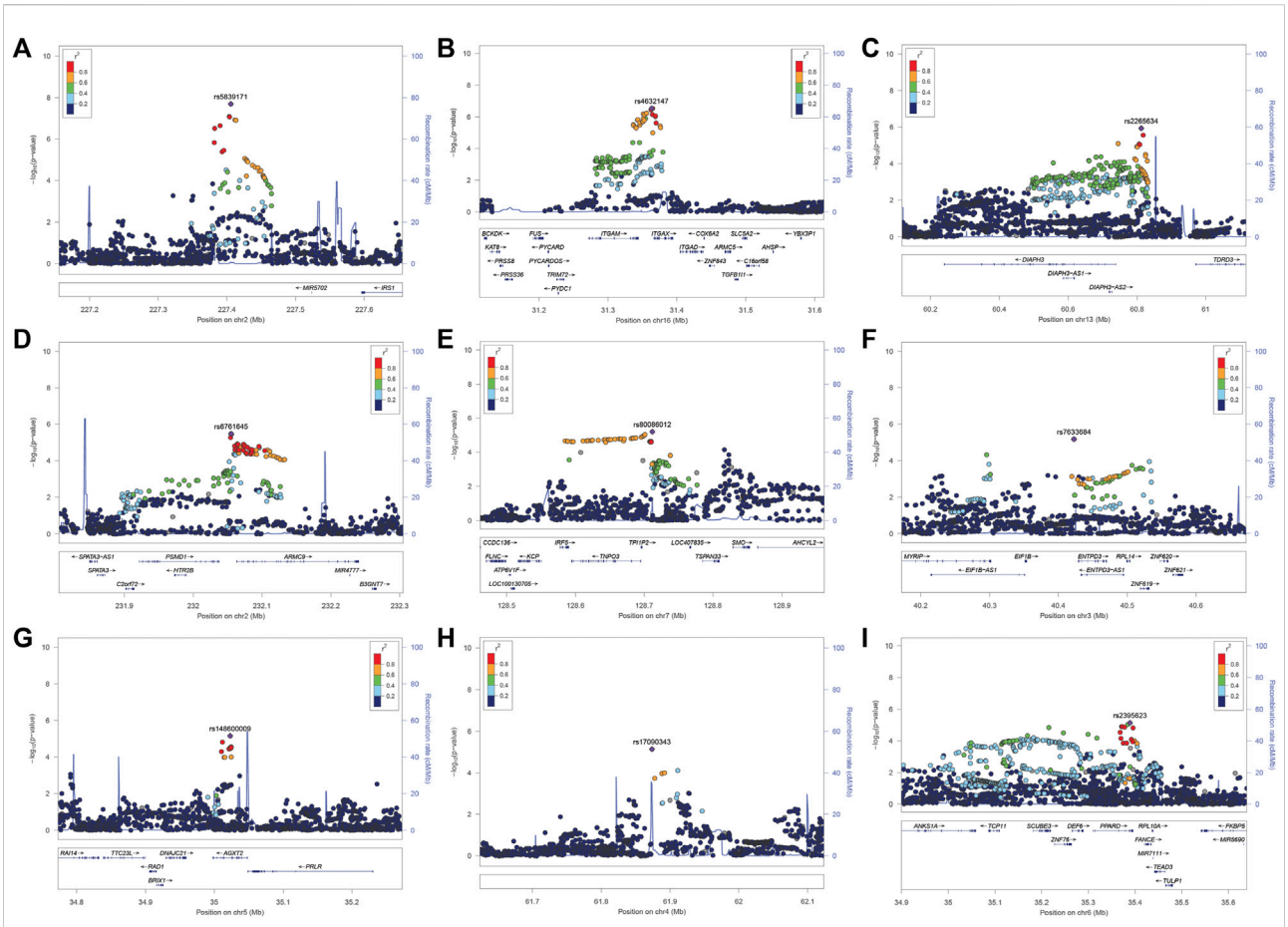


FIGURE 3
(A–I). LocusZoom plot near the *IRS1/miR-5702* locus, which reached genome-wide significance, and 8 suggestive loci.

identify independent associations at known loci (Table 2). The most significant SNPs at each locus are listed in Table 2. Among them, prominent ($P_{perm} < 5.0 \times 10^{-5}$) signals include *ITGAM-ITGAX* (rs71391210; $P_{perm} = 2.0 \times 10^{-7}$), *IRF5* (rs80086012; $P_{perm} = 3.5 \times 10^{-6}$), *DEF6-PPARD* (rs2395623; $P_{perm} = 4.70 \times 10^{-6}$), *XKR6* (rs2409660; $P_{perm} = 4.55 \times 10^{-5}$), *IRF1* (rs10586626; $P_{perm} = 5.43 \times 10^{-5}$), and *TYK2* (rs77389625; $P_{perm} = 1.91 \times 10^{-4}$). All replicated ($P_{perm} < 0.01$) SNPs and their nearest genes are shown in Table 2.

3.2.3 Minor allele frequency of associated SNPs and ethnicity

Allelic abundance—most often reported as minor allele frequency (MAF)—is an important measure both to estimate the effect of given SNPs on specific populations and more broadly to compare ancestry and population stratification between diverse populations. Using the 97 total replicated lead SNPs, we compared MAFs in our Egyptian cohort with those of Africans, Europeans, and Asians. We found a remarkable similarity between the MAFs of Egyptians and Europeans (R^2 = 93%)—i.e., the MAF variance in Egyptians was well explained by the MAF variance in Europeans. Conversely, correlation with Africans (R^2 = 51%) and Asians (R^2 = 59%) was markedly lower (Figure 4). Although surprising *prima facie*, these results were supported by our admixture analysis (Supplementary Figure S3).

3.2.4 Functional SNPs from some notable loci

Having identified several new SLE-associated loci, most of which have known roles in both immune homeostasis and other autoimmune disorders, we sought to better establish possible mechanisms for both the loci and the associated SNPs. For our nine novel loci, we first looked for significant ($P < 1 \times 10^{-5}$) eQTLs for both the top SNPs and the larger set of SNPs in linkage disequilibrium ($r^2 > 0.8$) with them. We found several significant ($3.5 \times 10^{-95} < P < 1 \times 10^{-2}$) eQTLs from our new loci, particularly from *ARMC9* (43 SNPs; 25 tissues), *IFLTD1* (5 SNPs; 4 tissues), *DIAPH3* (6 SNPs; 1 tissue), and *ENTPD3* (1 SNP; 10 tissues) (Figure 5A). *IFLTD1* is enriched in the brain, sex organs, and digestive

TABLE 2 Replication of 97 published loci at P_{perm}<0.01.

Chr	Published_SNP	PMID	Top_SNP	Nearest genes	r^2_pub_top	A1	F_A	F_U	A2	P_{uncorr}	P_{admix}	P_{perm}	#Sig_SNP	OR
16	rs34572943	28714469	rs71391210	ITGAM, ITGAX	0.40	T	0.38	0.284	C	8.97E-07	3.91E-07	2.00E-07	170	1.56
7	rs3757387	33272962	rs80086012	IRF5, TNPO3	0.06	T	0.166	0.11	C	3.00E-05	1.37E-05	3.50E-06	101	1.72
6	rs10807150	26808113	rs2395623	ANKS1A, PPARD, UHRF1BP1, DEF6	0.40	C	0.466	0.567	T	1.22E-06	3.65E-06	3.70E-06	205	0.66
8	rs7819602	28714469	rs2409660	AC011008.2, XKR6	0.09	A	0.235	0.168	G	5.77E-05	5.10E-05	4.55E-05	44	1.52
19	rs12461589	33272962	rs68013007	ANKRD27, PDCD5	0.01	T	0.203	0.144	A	4.62E-05	2.78E-05	4.69E-05	24	1.63
2	rs9630991	33536424	rs59207796	AC108047.1	0.56	C	0.206	0.273	CACATG	7.03E-05	1.10E-04	5.34E-05	110	0.65
5	rs2549002	33272962	5:131813034	IRF1	0.30	A	0.267	0.203	AAAG	1.15E-04	5.99E-05	5.43E-05	92	1.5
6	rs35789010	28714469	rs1321248	CARMIL1	0.01	C	0.2	0.267	T	1.32E-04	1.11E-04	6.97E-05	13	0.67
6	rs148314165	33272962	rs374184737	BTF3L4P3, LINC02528, TNFAIP3	0.04	T	0.476	0.569	TAA	1.20E-05	5.14E-05	7.48E-05	59	0.69
1	rs4844538	33272962	rs28584674	IKBKE, IL10, IL19, SRGAP2	0.00	C	0.081	0.129	T	3.17E-05	2.77E-05	1.42E-04	13	0.5
19	rs55882956	33272962	rs77389625	TYK2	mono	A	0.042	0.075	G	2.98E-04	2.78E-04	1.91E-04	16	0.44
8	rs16902895	33272962	rs6470627	LINC00824	0.63	G	0.848	0.898	T	1.63E-04	5.32E-04	2.92E-04	70	0.61
7	rs4598207	33272962	rs4917016	C7orf72, IKZF1	0.06	C	0.722	0.779	T	1.95E-04	2.76E-04	2.94E-04	5	0.66
1	rs2205960	33272962	rs34697014	LOC100506023, TNFSF4	0.00	G	0.251	0.312	A	6.16E-04	3.80E-04	3.77E-04	85	0.71
10	rs111447985	33272962	rs1008463	STN1	0.00	G	0.864	0.907	A	1.87E-04	4.15E-04	3.80E-04	9	0.57
5	rs2431697	33272962	rs1422980	MIR3142, MIR3142HG	0.03	T	0.368	0.3	C	4.82E-04	5.96E-04	4.06E-04	28	1.37
19	rs4801882	33272962	rs11672086	SIGLEC5	0.02	C	0.289	0.228	T	2.53E-04	1.46E-04	4.78E-04	55	1.46
13	rs1885889	33536424	rs11619751	AL136961.1, TM9SF2	0.47	A	0.136	0.094	G	1.04E-03	7.53E-04	5.24E-04	15	1.56
11	rs10896045	33272962	rs36089663	AP5B1, OVOL1	0.01	TA	0.296	0.237	T	5.24E-04	6.23E-04	5.85E-04	14	1.42
13	rs76725306	33536424	rs10675447	AL135901.1, RCBTB1	0.00	TTA	0.749	0.686	T	3.30E-04	3.21E-04	6.08E-04	104	1.45
11	rs77885959	33272962	rs4150658	GTF2H1	mono	G	0.144	0.113	A	1.08E-02	1.97E-03	6.36E-04	2	1.45
1	rs3806357	33272962	rs72744822	ELF3	0.00	A	0.136	0.096	G	1.25E-03	8.35E-04	6.46E-04	32	1.58
12	rs77465633	33272962	rs66480035	ATXN2	mono	C	0.463	0.524	T	6.80E-04	3.65E-04	6.85E-04	80	0.71
1	rs76107698	33272962	rs140778333	AL590385.2, FCGR2A, FCGR2C	0.00	T	0.022	0.049	C	8.21E-04	9.38E-04	6.99E-04	10	0.42
9	rs1887428	33272962	rs33973400	JAK2	0.00	C	0.371	0.441	CT	2.08E-04	2.14E-04	7.67E-04	17	0.71
17	rs2671655	33272962	rs140070508	LOC102724596	0.00	G	0.044	0.022	A	1.85E-03	2.81E-03	7.79E-04	30	2.16
22	rs4819670	33272962	rs5993014	USP18	0.25	A	0.3	0.242	G	9.45E-04	9.02E-04	8.56E-04	3	1.39
4	rs6841907	33272962	rs1129617	COQ2	0.21	A	0.225	0.284	G	4.16E-04	6.94E-04	8.60E-04	42	0.68
10	rs7097397	33272962	rs2620895	LRRC18, PCDH15, WDFY4	0.04	A	0.121	0.076	G	1.63E-04	8.47E-04	8.83E-04	3	1.74
8	rs2428	29625966	rs117234614	MFHAS1	0.03	C	0.064	0.043	T	7.76E-03	3.19E-03	9.56E-04	32	1.75

(Continued on following page)

TABLE 2 (*Continued*) Replication of 97 published loci at P_{perm}<0.01.

Chr	Published_SNP	PMID	Top_SNP	Nearest genes	r^2_pub_top	A1	F_A	F_U	A2	P_{uncorr}	P_{admix}	P_{perm}	#Sig_SNP	OR
2	rs3087243	33536424	rs11374410	CTLA4, ICOS	0.00	GA	0.127	0.088	G	1.03E-03	6.19E-04	1.01E-03	19	1.6
1	rs1547624	33536424	rs10921148	AL390957.1	0.01	T	0.195	0.147	C	4.53E-04	3.74E-04	1.02E-03	6	1.56
12	rs6539078	33272962	rs703619	AC084364.4, LOC105369945	0.00	A	0.373	0.439	G	1.42E-03	1.63E-03	1.09E-03	8	0.76
17	rs2286672	26502338	rs3786042	PLD2	0.98	T	0.104	0.068	C	3.10E-04	3.36E-04	1.10E-03	4	1.87
1	rs1780813	29848360	rs142583842	SMYD3	0.65	ATAGC	0.031	0.057	A	2.66E-03	2.09E-03	1.16E-03	52	0.5
5	rs2421184	33272962	rs17056704	LINC01845	0.02	C	0.217	0.169	G	1.48E-03	1.95E-03	1.19E-03	38	1.44
16	rs11117432	33272962	rs11117444	AC092723, IRF8	0.00	C	0.219	0.272	G	9.96E-04	6.20E-04	1.22E-03	25	0.69
12	rs2540119	33272962	rs142451408	PARP11	0.09	A	0.504	0.562	AAAG	2.40E-03	1.65E-03	1.23E-03	16	0.76
15	rs35985016	33272962	rs1455854	LRRK1	0.00	T	0.378	0.319	C	1.01E-03	6.47E-04	1.35E-03	24	1.38
14	rs12148050	33536424	rs4900554	TRAF3	0.02	A	0.122	0.153	G	1.42E-02	7.56E-03	1.36E-03	3	0.7
19	rs10419308	33536424	rs7259964	AC010327, TMEM86B	0.00	G	0.133	0.088	T	3.91E-04	1.08E-03	1.37E-03	3	1.63
6	rs597325	33272962	rs13209535	BACH2	0.00	G	0.122	0.091	A	9.83E-03	1.99E-03	1.56E-03	14	1.46
3	rs7637844	33272962	rs6779548	LINC00870	0.03	G	0.656	0.718	A	1.66E-03	1.50E-03	1.56E-03	12	0.76
6	rs36014129	28714469	rs77341667	H2AC3P, H2BP5	0.01	T	0.032	0.056	C	5.68E-03	3.49E-03	1.57E-03	2	0.53
12	rs200521476	33272962	rs11610045	FBRSL1	0.11	A	0.361	0.428	G	1.12E-03	2.18E-03	1.63E-03	34	0.75
2	rs11889341	33272962	rs11677408	STAT4	0.02	T	0.063	0.04	C	4.33E-03	2.66E-03	1.68E-03	3	1.83
1	rs12093154	33536424	rs1240748	C1QTNF12	0.08	T	0.708	0.657	C	7.85E-03	1.92E-03	1.72E-03	9	1.29
13	rs57141708	33272962	rs9566681	ELF1	0.20	T	0.11	0.077	C	2.60E-03	4.37E-03	1.80E-03	1	1.61
11	rs2785198	33272962	11:35203468	LOC100507144, PDHX	0.01	TTA	0.418	0.476	T	2.51E-03	1.84E-03	1.90E-03	11	0.76
12	rs11059928	33272962	rs5801815	SLC15A4	0.00	C	0.352	0.406	CA	2.32E-03	2.92E-03	2.08E-03	4	0.73
18	rs1788097	33272962	rs34594414	CD226	0.24	T	0.381	0.322	A	6.89E-04	1.21E-03	2.25E-03	4	1.41
10	rs77448389	33272962	rs2246268	ANKRD16	0.02	C	0.065	0.04	A	1.82E-03	2.32E-03	2.42E-03	7	1.96
12	rs4622329	33272962	rs144176212	DRAM1	0.04	C	0.707	0.763	CTTTT	6.63E-04	4.10E-04	2.49E-03	23	0.7
11	rs4930642	33272962	rs2924520	TPCN2	0.04	G	0.378	0.442	A	1.17E-03	1.82E-03	2.53E-03	30	0.74
8	rs17374162	33272962	rs2099950	AS1, MSC	0.04	C	0.203	0.159	G	4.58E-03	3.82E-03	2.55E-03	13	1.37
19	rs7251	32719713	rs8103298	IRF3	0.02	C	0.056	0.087	T	5.48E-03	3.73E-03	2.58E-03	4	0.63
10	rs10823829	33272962	rs7909048	CDH23	0.05	T	0.066	0.038	G	1.43E-03	3.24E-03	2.74E-03	7	1.86
22	rs4821116	33272962	rs4820091	CCDC116, UBE2L3, YDJC	0.57	G	0.349	0.289	T	1.53E-03	2.09E-03	2.77E-03	67	1.33
4	rs13101828	33272962	rs41286651	DGKQ	0.04	C	0.043	0.022	T	5.18E-03	5.06E-03	3.02E-03	1	1.94
3	rs6762714	27399966	rs1152846	LPP	0.19	C	0.664	0.718	T	5.40E-03	4.28E-03	3.03E-03	13	0.78

(Continued on following page)

TABLE 2 (*Continued*) Replication of 97 published loci at P_{perm}<0.01.

Chr	Published_SNP	PMID	Top_SNP	Nearest genes	r^2_pub_top	A1	F_A	F_U	A2	P_{uncorr}	P_{admix}	P_{perm}	#Sig_SNP	OR
10	rs58164562	33272962	rs2031130	BBIP1	0.01	G	0.588	0.644	T	6.03E-03	4.49E-03	3.05E-03	4	0.79
1	rs2476601	26502338	rs6663888	PHTF1, PTPN22, RSBN1	0.00	A	0.035	0.051	C	2.71E-02	5.90E-03	3.07E-03	3	0.56
5	rs10036748	33272962	rs960709	TNIP1	0.96	G	0.425	0.36	A	1.21E-03	2.53E-03	3.15E-03	6	1.32
14	rs2819426	33272962	rs33925946	AHNAK2, AHNAK2, PLD4	0.04	T	0.176	0.215	G	9.39E-03	4.64E-03	3.23E-03	5	0.73
16	rs11288784	33272962	rs11076521	HEATR3	0.11	T	0.628	0.582	C	8.48E-03	3.79E-03	3.46E-03	10	1.3
2	rs13385731	33272962	rs4569454	RASGRP3	0.00	G	0.628	0.69	A	9.61E-04	1.02E-03	3.80E-03	2	0.74
1	rs13306575	33272962	rs17849502	NCF2, NMNAT2, SMG7	mono	T	0.072	0.044	G	5.72E-03	5.81E-03	3.81E-03	2	1.61
17	rs8072449	28714469	rs147427721	AC011933.4, GRB2, SLC25A19	0.15	T	0.876	0.849	TTC	4.29E-02	1.06E-02	3.90E-03	6	1.31
1	rs116785379	33272962	1:157081831	ETV3	0.00	G	0.075	0.05	C	4.30E-03	5.79E-03	3.92E-03	1	1.73
6	rs9488914	33272962	rs1204826	DSE	0.29	G	0.665	0.708	A	1.88E-02	1.03E-02	4.12E-03	4	0.8
18	rs118075465	33272962	rs34787266	LOC284241	0.00	T	0.094	0.129	C	9.59E-03	5.27E-03	4.38E-03	10	0.7
3	rs564976	26502338	rs7652547	AS1, IL12A	0.01	C	0.198	0.247	T	6.09E-03	5.10E-03	4.46E-03	7	0.76
4	rs231694	33272962	rs116412781	FAM193A, TNIP2	0.15	T	0.078	0.047	C	9.11E-04	6.70E-04	4.53E-03	3	1.84
16	rs11376510	33272962	rs889791	MAFTRR	0.00	T	0.098	0.127	G	1.37E-02	8.79E-03	4.59E-03	5	0.68
17	rs35966917	33272962	rs79185281	TNFRSF13B	0.10	C	0.542	0.589	CT	1.02E-02	1.26E-02	4.61E-03	0	0.78
6	rs9322454	33272962	rs12199124	IPCEF1	0.14	A	0.066	0.042	G	7.21E-03	7.41E-03	4.69E-03	0	1.67
1	rs3795310	33536424	rs72635735	RERE	0.00	C	0.052	0.032	T	6.19E-03	8.01E-03	4.71E-03	1	1.89
4	rs2855772	29494758	rs2646326	KIT	0.00	A	0.068	0.046	G	1.44E-02	8.59E-03	5.32E-03	0	1.6
11	rs9736939	33272962	rs12792061	AP001122.1, ETS1, LINC02098	0.02	A	0.112	0.152	G	4.76E-03	7.73E-03	5.82E-03	4	0.69
1	rs11264750	33536424	rs10796979	FCRL5	0.20	C	0.854	0.811	A	6.49E-03	4.26E-03	6.20E-03	16	1.37
3	rs9852465	28714469	rs7638793	AC116036.2, PDHB, PXK	0.05	C	0.274	0.224	T	2.77E-03	4.33E-03	6.22E-03	27	1.36
5	rs6871748	33536424	rs79041667	AC112204.3, IL7R	0.01	G	0.193	0.149	A	3.85E-03	3.51E-03	6.28E-03	8	1.39
1	rs28411034	33536424	rs34755028	MTF1	0.06	C	0.671	0.721	CA	6.10E-03	7.74E-03	6.75E-03	1	0.77
3	rs144104218	33272962	rs20568	CD80, TIMMDC1, TMEM39A	0.01	T	0.098	0.068	C	1.07E-02	7.91E-03	6.82E-03	3	1.47
1	rs6702599	33536424	rs35189936	IL12RB2	0.00	A	0.076	0.112	AT	2.41E-03	3.92E-03	7.25E-03	19	0.61
8	rs2736332	33272962	8:11346080	AF131216.5, BLK	0.13	G	0.117	0.083	T	1.93E-03	2.56E-03	7.37E-03	4	1.62
10	rs7902146	33272962	rs76465877	ARID5B	0.00	A	0.048	0.028	G	6.67E-03	7.34E-03	7.41E-03	1	1.84
17	rs61759532	33272962	rs7214863	AC026954.1, ACAP1	0.00	T	0.629	0.591	C	3.21E-02	1.16E-02	7.52E-03	1	1.23
7	rs117026326	33272962	rs10256306	GTF2IRD1, LOC101926943	0.01	A	0.072	0.049	G	1.95E-02	9.88E-03	7.53E-03	1	1.5
14	rs911263	28714469	rs17828548	RAD51B	0.01	T	0.047	0.07	C	1.03E-02	1.21E-02	7.90E-03	0	0.59

(Continued on following page)

TABLE 2 (*Continued*) Replication of 97 published loci at P_{perm}<0.01.

Chr	Published_SNP	PMID	Top_SNP	Nearest genes	r^2_pub_top	A1	F_A	F_U	A2	P_{uncorr}	P_{admix}	P_{perm}	#Sig_SNP	OR
4	rs58107865	33272962	rs10516550	LEF1	mono	T	0.047	0.069	C	8.88E-03	6.80E-03	7.92E-03	2	0.55
5	rs7725218	33272962	rs72715511	TERT	0.00	T	0.223	0.186	C	1.92E-02	6.28E-03	8.22E-03	1	1.3
15	rs8023715	24871463	rs79487017	LINC02253, RN7SKP181	0.00	A	0.03	0.058	G	2.08E-03	3.92E-03	8.43E-03	3	0.49
12	rs4251697	33272962	rs116902872	CDKN1B, CREBL2, GPR19	0.00	A	0.025	0.045	G	1.25E-02	1.07E-02	8.74E-03	0	0.54
2	rs2381401	33536424	rs10048784	ARHGAP15	0.01	C	0.707	0.662	T	2.36E-02	1.33E-02	8.99E-03	3	1.22
4	rs10018951	29494758	4:184474355	TRAPPC11	0.00	T	0.913	0.888	C	2.43E-02	2.48E-02	9.18E-03	0	1.45
4	rs4643809	33272962	rs79879350	BANK1	0.09	T	0.046	0.07	C	5.69E-03	2.56E-03	9.35E-03	8	0.53
8	rs2953898	28714469	rs115578289	RPS20	0.01	C	0.044	0.026	T	1.53E-02	1.24E-02	9.37E-03	1	1.78

mono = Monomorphic SNP

tract (https://www.proteinatlas.org), and most eQTL signals came from these tissues. *ARMC9* is very strongly expressed in natural killer (NK) cells (https://www.proteinatlas.org) and melanocytes (Otani et al., 2006); interestingly, the strongest eQTL signals arose from vascular endothelium and thyroid. Known loci also contained many eQTL SNPs; for example, *IRF1* (58 SNPs; 23 tissues) showed particularly strong eQTL signals. *IRF1* is quite immune cell-specific, and accordingly, the strongest eQTL signals came from blood (Figure 5B).

4 Discussion

Our study represents the first lupus GWAS in North Africans, an underrepresented population in human genetic studies. Previous studies indicated that Egyptians are highly admixed, with substantial ancestry arising from the Middle East, Europe, and Africa—with the extent of European and Caucasus hunter-gatherer (Lazaridis et al., 2016) contribution being higher, and the African contribution being lower, than might be expected from geography. Greater European/Middle Eastern than African ancestry is also supported by whole-genome sequencing (ElHefnawi et al., 2021). Importantly, modern Egyptians are quite divergent from ancient populations. The inclusion of more diverse participants in genomics studies has been shown to significantly increase fine-mapping resolution (Zaitlen et al., 2010; Asimit et al., 2016). Further studies in Egyptians (and other North African and Middle Eastern populations) could potentially reveal more genetic associations, with implications for understanding, diagnosing, and treating disease for Egyptians and all humans.

In this study, we identified nine novel loci; one is genome-wide significant for lupus susceptibility and eight are suggestive. Among 179 known non-HLA SLE susceptibility loci, 97 were successfully replicated—indicating that most lupus susceptibility in Egyptians is ancestry-independent. We also found several loci not previously reported in any populations. Intriguingly, the novel loci were substantially enriched in proteins involved in cellular proliferation and cancer risk. Specifically, Insulin receptor substrate 1 (*IRS1*), the microRNA miR-5702, Diaphanous homolog 1 (*DIAPH3*), Armadillo repeat containing 9 (*ARMC9*, *a.k.a.* KU-MEL-1), Intermediate filament tail domain containing 1 (*IFLTD1*, *a.k.a. LMNTD1*), immune adaptor protein NCK2 (Labelle-Côté et al., 2011), and Ectonucleoside triphosphate diphosphohydrolase 3 (*ENTPD3*) (Wang J. et al., 2021) all have well-documented roles in cellular proliferation and tumor invasion. Proliferation factors play a prominent role in lupus, particularly in invasion of specific organs in sub-clinical phenotypes, most notably lupus nephritis (Balomenos et al., 2000; Treamtrakanpon et al., 2012).

The novel and replicated loci in this study share other features in addition to association with cellular proliferation—for instance, regulation of and by microRNAs, another hallmark of SLE (Dai et al., 2007; Carlsen et al., 2013). Notably, both *IRS1* and *ARMC9* (along with *LRIG2*, *PSPH*, and *SKP2*) are direct targets down-regulated by miRNA-150 (Zhang et al., 2021). miRNA-150 is a critical immune regulator, for instance driving expression of the inflammatory receptor TREM-1 (Triggering receptor expressed on myeloid cells 1) in dendritic cells and contributing to lupus-like symptoms in mice (Gao et al., 2017). More recently, miRNA-150 (along with miRNA-148b) has been proposed as a specific biomarker for lupus nephritis (Alkhateeb and Altamemi, 2021). Two of our novel risk loci even interact at the protein level: NCK2 directly binds to IRS1 to

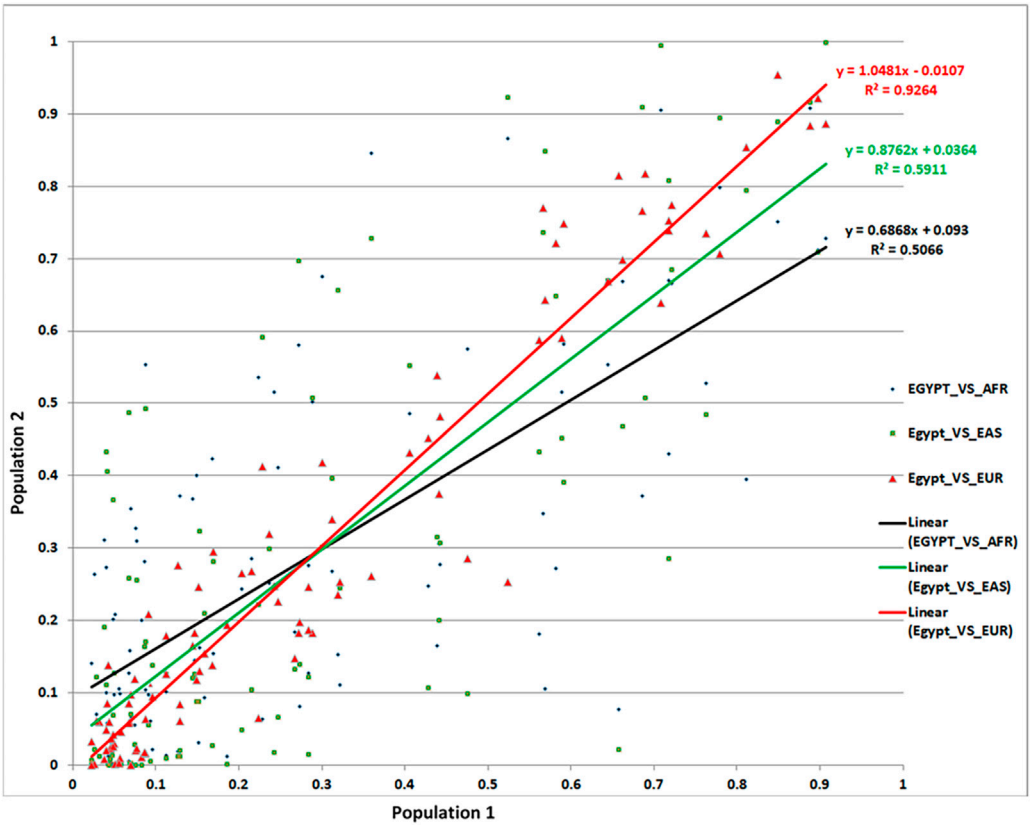


FIGURE 4
Comparison of MAFs for the associated SNPs from our Egyptian controls and 1,000 Genomes data.

create an intracellular signal transduction mechanism for signaling through insulin and insulin-like growth factor receptors (Tu et al., 1998; Tu et al., 2001). We performed a sequence search with our novel risk miRNA *miR548AG1* and discovered ten closely related sequences, all of them in introns or UTRs of immune-critical genes (Supplementary Table S5) –many overlapping enhancers and/or transcription factor binding sites. Furthermore, five of these ten hits are ESTs of ~90 bp–100 bp and almost certainly constitute previously unannotated members of the miR-548 family. The clear immune association of all ten closely related sequences further confirms the relevance of our SLE risk locus *miR548AG1*.

Our study very carefully controlled for admixture in estimating effect size and significance level. This was a critical step, as multiple loci became more significant or less significant once controlled for admixture (Tables 1, 2; Supplementary Table S1). The most substantive change was at the novel locus *IRS1/miR-5702,* where the top SNP rs5839171 achieved genome-wide significance after admixture correction ($P_{uncorrected} = 1.70 \times 10^{-7}$, $P_{corrected} = 1.98 \times 10^{-8}$). This remains significant ($P_{permuted} = 3.3 \times 10^{-8}$) after 10^9 permutations. This emphasizes the utility of admixed populations, like Egyptians, in studying novel gene discovery.

To assess the replication of previously known loci [Table 1 from (Ha et al., 2022)], we implemented a very large (10^7–10^9 combinations) ancestry-adjusted permutation analysis on our data. Critically, permutation analysis controls for low sample-size effects, as all permutations have identical n as the proposal—comparisons are between permutations with identical statistical power *a priori* to produce an association. Using the ancestry-corrected permutation-based analysis, we set threshold for replication at $p < 0.01$. Here we further explain our reasoning behind the robustness of our permutation analysis and some of its findings.

First, a recent large-scale (n > 208,000) SLE GWAS meta-analysis on East-Asian populations identified rs956237 (GWAS lead variant; intron of *LEF1*, Lymphoid enhancer-binding factor 1—a leukocyte transcription factor) at 4q25 as a novel susceptibility locus (Yin et al., 2020; $p < 4.0 \times 10^{-11}$). Reassuringly, a companion transcriptome-wide association study (TWAS) using data from Asian samples also strongly indicated *LEF1* as the primary gene underlying 4q25 association (Yin et al., 2020; $p < 1.3 \times 10^{-10}$). However, our primary reference (Table 1 from Ha et al., 2022) nominated rs58107865 instead of rs956237 as lead SNP. In our study, we

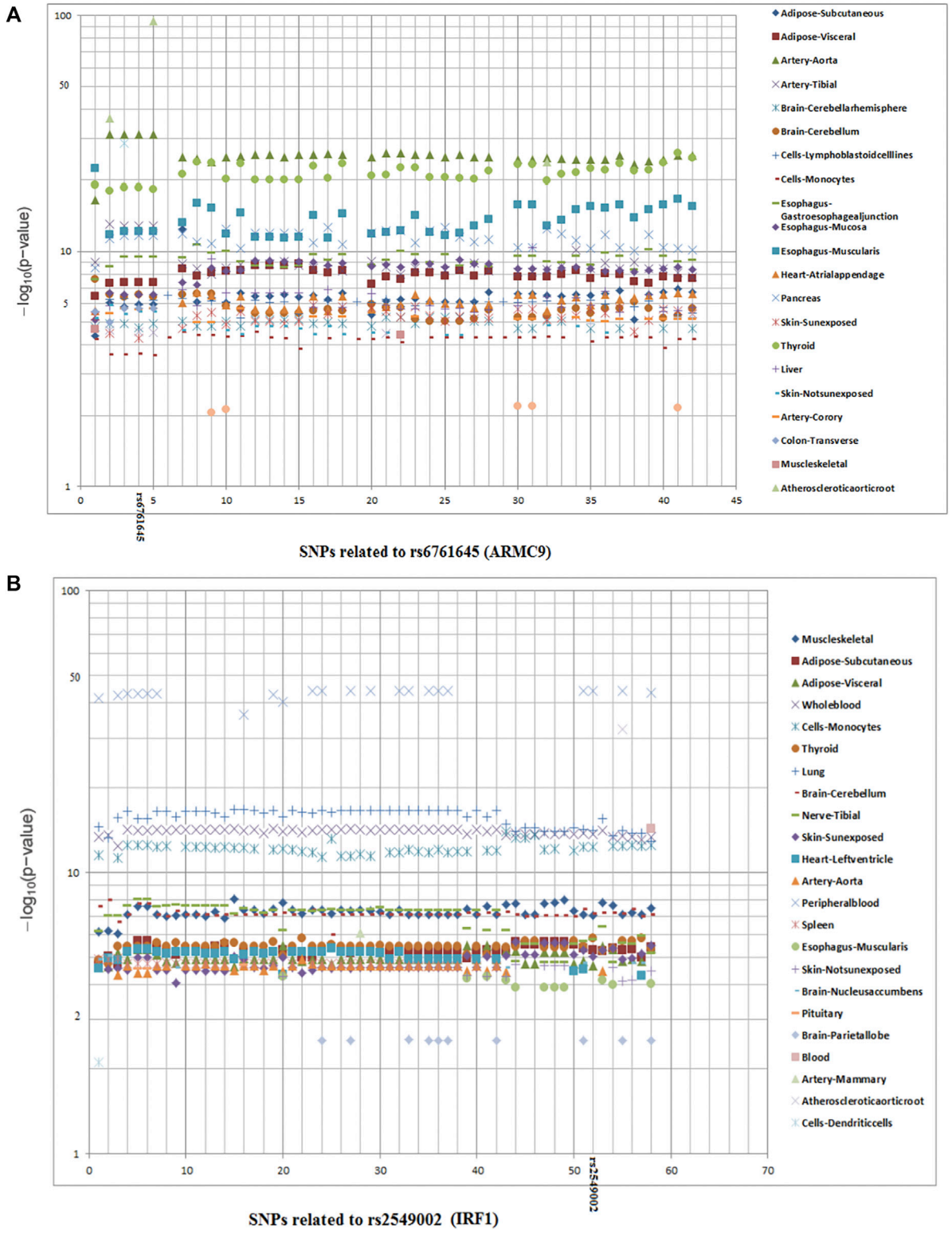


FIGURE 5
(A,B). eQTL plots for the related SNPs (r^2 > 0.8 with top SNPs) in two notable loci.

found that rs58107865 does not exist in the African/Egyptian population. Instead, we discovered an association with another intronic *LEF1* SNP (rs10516550, p = 0.008), common in African/European/Hispanic populations (but not Asian). Of course, this SNP needs to be independently replicated in a population with similar ethnic background. But we believe that it's appropriate to cautiously interpret this in the replication category.

Second, *NCF2* is a very well-known SLE locus, where multiple studies (Cunninghame Graham et al., 2011; Yin et al., 2020) including ours (Kim-Howard et al., 2014) have reported non-coding and at least 3 independent ethnicity-specific coding variants within *NCF2* passing genome-wide significance in Asian/Hispanic (rs13306575), European/African-American (rs17849502) and African-American (rs35937854) populations. This locus was represented by rs13306575 in our reference data (Table 1 from Ha et al., 2022). In our data, we found evidence of association (ancestry-adjusted permutation p = 0.004) with nearby (< 150 bp) rs17849502. Thus, we have found essentially the same result as the prior studies; however, this real association falls below the strict threshold set by the Bonferroni correction (p = 0.05/179 = 0.00027). Our non-parametric permutation analysis, however, detects this locus as a significant association, which we believe is supported by the preponderance of evidence.

Although the sample sizes in the current study are not overly large, we had the power to attain multiple noteworthy discoveries for several reasons: 1) The highly admixed Egyptian population gave us the power to discover nine novel loci, one of which achieved genome-wide significance. 2) Our data attained a remarkable replication rate (97 out of 179 loci; 54%)—increasing confidence in both known and novel loci, and supporting the notion that despite different genetic backgrounds, underlying loci and risk alleles are broadly conserved across ethnicities (Marigorta and Navarro, 2013). Over 80% of our replicated loci arose from GWASs on Europeans, indicating broad concordance of genetic risk. However, it is important to note that the studies included in Ha et al. (2022) were performed on populations with large sample sizes and statistical power. Egyptians have a stronger contribution from Middle Eastern ancestries than European (Wohlers et al., 2020), and their population history and admixture proportions are different. Moreover, our sample size was relatively small. Future studies will boost sample sizes and further address the other aspects. 3) We found several common pathways shared by the known and new loci, increasing the robustness of the observations, and laying out the foundation for further experimental studies to establish a clear mechanistic basis for these loci contributing to SLE risk and progression.

However, we also acknowledge that lack of an independent replication, especially for the novel associations, is a limitation of this study. Therefore, in addition to strengthening the observations and establishing mechanisms, further studies with larger sample sizes will clearly delineate the ancestry-independent and ancestry-specific components of SLE association.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the Oklahoma Medical Research Foundation (IRB approval number 14–16), and Mansoura University (Egypt) Ethics committee/IRB: MFM-IRB code number R/16.04.81.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving human participants were reviewed and approved by Institutional Review Board of the Oklahoma Medical Research Foundation And Mansoura University Ethics Committee/Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

Author contributions

Conceptualization, GR and SN; Data curation, AAE, CS, LL and MTH; Formal analysis, CS and SN; Funding acquisition, SI and SN; Investigation, ZT, MA, WA, IE-S, ME, AnA, SH, NS, AH and ENA; Methodology, CS, LL and SN; Resources, NMK, MB, MTH, MAH, MNS, EE, ZT, MA, WA, IE-S, ME, AnA, SH, NS, AS, AmE, AmH, KR, HE-T, AY, FE, AAH, AsA, EmA, DF, SE-B, ME, ENA, and SN; Software, CS, LL, NK, MB and SN; Supervision, NK, SI, GR and SN; Validation, AAE, CS, MAH, MNS and SI; Visualization, AAE, SI, GR and SN; Writing–review and editing, AAE, LL, CS, SE-K, GR and SN. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the University of Lübeck and by the National Institutes of Health, grant numbers MD007909 and AR060366.

Acknowledgments

We thank all patients with SLE and healthy controls who participated in this study. We also thank the research assistants, coordinators, and physicians who helped in the recruitment of subjects for this project.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.948505/full#supplementary-material.

References

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19 (9), 1655–1664. doi:10.1101/gr.094052.109

Alkhateeb, D. M., and Altamemi, A. (2021). Circulating microrna (148b and 150) as potential biomarker in iga nephropathy and lupus nephritis. *Ann. Romanian Soc. Cell Biol.* 25 (6), 6294–6309.

Asimit, J. L., Hatzikotoulas, K., McCarthy, M., Morris, A. P., and Zeggini, E. (2016). Trans-ethnic study design approaches for fine-mapping. *Eur. J. Hum. Genet.* 24 (9), 1330–1336. doi:10.1038/ejhg.2016.1

Balomenos, D., Martín-Caballero, J., García, M. I., Prieto, I., Flores, J. M., Serrano, M., et al. (2000). The cell cycle inhibitor p21 controls T-cell proliferation and sex-linked lupus development. *Nat. Med.* 6 (2), 171–176. doi:10.1038/72272

Bejaoui, Y., Witte, M., Abdelhady, M., Eldarouti, M., Abdallah, N. M. A., Elghzaly, A. A., et al. (2019). Genome-wide association study of psoriasis in an Egyptian population. *Exp. Dermatol.* 28 (5), 623–627. doi:10.1111/exd.13926

Ben-Eghan, C., Sun, R., Hleap, J. S., Diaz-Papkovich, A., Munter, H. M., Grant, A. V., et al. (2020). Don't ignore genetic data from minority populations. *Nature* 585 (7824), 184–186. doi:10.1038/d41586-020-02547-3

Bentham, J., Morris, D. L., Cunninghame Graham, D. S., Pinder, C. L., Tombleson, P., Behrens, T. W., et al. (2015). Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat. Genet.* 47 (12), 1457–1464. doi:10.1038/ng.3434

Block, S. R. (2006). A brief history of twins. *Lupus* 15 (2), 61–64. doi:10.1191/0961203306lu2263ed

Cakmak Genc, G., Dursun, A., Karakas Celik, S., Calik, M., Kokturk, F., and Piskin, I. E. (2018). IL28B, IL29 and micro-RNA 548 in subacute sclerosing panencephalitis as a rare disease. *Gene* 678, 73–78. doi:10.1016/j.gene.2018.07.062

Calin, G. A., and Croce, C. M. (2006). MicroRNA-cancer connection: The beginning of a new tale. *Cancer Res.* 66 (15), 7390–7394. doi:10.1158/0008-5472.Can-06-0800

Cannons, J. L., Zhao, F., and Schwartzberg, P. L. (2013). Ncking BCR-mediated PI3K activation. *EMBO Rep.* 14 (10), 852–853. doi:10.1038/embor.2013.133

Carlsen, A. L., Schetter, A. J., Nielsen, C. T., Lood, C., Knudsen, S., Voss, A., et al. (2013). Circulating MicroRNA expression profiles associated with systemic lupus erythematosus. *Arthritis Rheum.* 65 (5), 1324–1334. doi:10.1002/art.37890

Carter, E. E., Barr, S. G., and Clarke, A. E. (2016). The global burden of SLE: Prevalence, health disparities and socioeconomic impact. *Nat. Rev. Rheumatol.* 12 (10), 605–620. doi:10.1038/nrrheum.2016.137

Chen, L., Ge, B., Casale, F. P., Vasquez, L., Kwan, T., Garrido-Martin, D., et al. (2016). Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* 167 (5), 1398–1414. doi:10.1016/j.cell.2016.10.026

Cohen, J. (1994). The Earth is round (p < .05). *Am. Psychol.* 49, 997–1003. doi:10.1037/0003-066X.49.12.997

Cohen, J. (1990). Things I have learned (so far). *Am. Psychol.* 45, 1304–1312. doi:10.1037/0003-066X.45.12.1304

Consortium, G. T. (2013). The genotype-tissue expression (GTEx) project. *Nat. Genet.* 45 (6), 580–585. doi:10.1038/ng.2653

Cortes, A., and Brown, M. A. (2011). Promise and pitfalls of the immunochip. *Arthritis Res. Ther.* 13 (1), 101. doi:10.1186/ar3204

Cummins, J. M., He, Y., Leary, R. J., Pagliarini, R., Diaz, L. A., Sjoblom, T., et al. (2006). The colorectal microRNAome. *Proc. Natl. Acad. Sci. U. S. A.* 103 (10), 3687–3692. doi:10.1073/pnas.0511155103

Cunninghame Graham, D. S., Morris, D. L., Bhangale, T. R., Criswell, L. A., Syvanen, A. C., Ronnblom, L., et al. (2011). Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with systemic lupus erythematosus. *PLoS Genet.* 7 (10), e1002341. doi:10.1371/journal.pgen.1002341

Dai, Y., Huang, Y. S., Tang, M., Lv, T. Y., Hu, C. X., Tan, Y. H., et al. (2007). Microarray analysis of microRNA expression in peripheral blood cells of systemic lupus erythematosus patients. *Lupus* 16 (12), 939–946. doi:10.1177/0961203307084158

Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., et al. (2016). Next-generation genotype imputation service and methods. *Nat. Genet.* 48 (10), 1284–1287. doi:10.1038/ng.3656

Deapen, D., Escalante, A., Weinrib, L., Horwitz, D., Bachman, B., Roy-Burman, P., et al. (1992). A revised estimate of twin concordance in systemic lupus erythematosus. *Arthritis Rheum.* 35 (3), 311–318. doi:10.1002/art.1780350310

Dehne, N., and Brüne, B. (2009). HIF-1 in the inflammatory microenvironment. *Exp. Cell Res.* 315 (11), 1791–1797. doi:10.1016/j.yexcr.2009.03.019

Elghzaly, A. A., Metwally, S. S., El-Chennawi, F. A., Elgayaar, M. A., Mosaad, Y. M., El-Toraby, E. E., et al. (2015). IRF5, PTPN22, CD28, IL2RA, KIF5A, BLK and TNFAIP3 genes polymorphisms and lupus susceptibility in a cohort from the Egypt Delta; relation to other ethnic groups. *Hum. Immunol.* 76 (7), 525–531. Epub 2015 Jun 17. doi:10.1016/j.humimm.2015.06.001

ElHefnawi, M., Hegazy, E., Elfiky, A., Jeon, Y., Jeon, S., Bhak, J., et al. (2021). Complete genome sequence and bioinformatics analysis of nine Egyptian females with clinical information from different geographic regions in Egypt. *Gene* 769, 145237. doi:10.1016/j.gene.2020.145237

Fadista, J., Manning, A. K., Florez, J. C., and Groop, L. (2016). The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur. J. Hum. Genet.* 24 (8), 1202–1205. doi:10.1038/ejhg.2015.269

Fukunaga, T., and Hamada, M. (2017). RIblast: An ultrafast RNA-RNA interaction prediction system based on a seed-and-extension approach. *Bioinformatics* 33 (17), 2666–2674. doi:10.1093/bioinformatics/btx287

Fukunaga, T., Iwakiri, J., Ono, Y., and Hamada, M. (2019). LncRRIsearch: A web server for lncRNA-RNA interaction prediction integrated with tissue-specific expression and subcellular localization data. *Front. Genet.* 10, 462. doi:10.3389/fgene.2019.00462

Gamazon, E. R., Zhang, W., Konkashbaev, A., Duan, S., Kistner, E. O., Nicolae, D. L., et al. (2009). Scan: SNP and copy number annotation. *Bioinformatics* 26 (2), 259–262. doi:10.1093/bioinformatics/btp644

Gao, S., Yuan, L., Wang, Y., and Hua, C. (2017). Enhanced expression of TREM-1 in splenic cDCs in lupus prone mice and it was modulated by miRNA-150. *Mol. Immunol.* 81, 127–134. doi:10.1016/j.molimm.2016.12.006

Gao, X. (2011). Multiple testing corrections for imputed SNPs. *Genet. Epidemiol.* 35 (3), 154–158. doi:10.1002/gepi.20563

Goulielmos, G. N., Zervou, M. I., Vazgiourakis, V. M., Ghodke-Puranik, Y., Garyfallos, A., and Niewold, T. B. (2018). The genetics and molecular pathogenesis of systemic lupus erythematosus (SLE) in populations of different ancestry. *Gene* 668, 59–72. doi:10.1016/j.gene.2018.05.041

Guan, T., Dominguez, C. X., Amezquita, R. A., Laidlaw, B. J., Cheng, J., Henao-Mejia, J., et al. (2018). ZEB1, ZEB2, and the miR-200 family form a counterregulatory network to regulate CD8+ T cell fates. *J. Exp. Med.* 215 (4), 1153–1168. doi:10.1084/jem.20171352

Ha, E., Bae, S. C., and Kim, K. (2022). Recent advances in understanding the genetic basis of systemic lupus erythematosus. *Semin. Immunopathol.* 44 (1), 29–46. doi:10.1007/s00281-021-00900-w

Hochberg, M. C. (1997). Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.* 40 (9), 1725. doi:10.1002/art.1780400928

Jima, D. D., Zhang, J., Jacobs, C., Richards, K. L., Dunphy, C. H., Choi, W. W. L., et al. (2010). Deep sequencing of the small RNA transcriptome of normal and malignant human B cells identifies hundreds of novel microRNAs. *Blood* 116 (23), e118–e127. doi:10.1182/blood-2010-05-285403

Johnson, J. L., and Abecasis, G. R. (2017). *GAS power calculator: Web-based power calculator for genetic association studies*. Cold Spring Harbor, NY: bioRxiv, 164343. doi:10.1101/164343

Johnson, R. C., Nelson, G. W., Troyer, J. L., Lautenberger, J. A., Kessing, B. D., Winkler, C. A., et al. (2010). Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genomics* 11 (1), 724. doi:10.1186/1471-2164-11-724

Julia, A., Lopez-Longo, F. J., Perez Venegas, J. J., Bonas-Guarch, S., Olive, A., Andreu, J. L., et al. (2018). Genome-wide association study meta-analysis identifies five new loci for systemic lupus erythematosus. *Arthritis Res. Ther.* 20 (1), 100. doi:10.1186/s13075-018-1604-1

Ke, H., Zhao, L., Feng, X., Xu, H., Zou, L., Yang, Q., et al. (2016). NEAT1 is required for survival of breast cancer cells through FUS and miR-548. *Gene Regul. Syst. Bio.GRSB* 10s1, 11–17. doi:10.4137/grsb.S29414

Kim-Howard, X., Sun, C., Molineros, J. E., Maiti, A. K., Chandru, H., Adler, A., et al. (2014). Allelic heterogeneity in NCF2 associated with systemic lupus erythematosus (SLE) susceptibility across four ethnic populations. *Hum. Mol. Genet.* 23 (6), 1656–1668. doi:10.1093/hmg/ddt532

Kiniwa, Y., Fujita, T., Akada, M., Ito, K., Shofuda, T., Suzuki, Y., et al. (2001). Tumor antigens isolated from a patient with vitiligo and T-cell-infiltrated melanoma. *Cancer Res.* 61 (21), 7900–7907.

Kuo, C. F., Grainge, M. J., Valdes, A. M., See, L. C., Luo, S. F., Yu, K. H., et al. (2015). Familial aggregation of systemic lupus erythematosus and coaggregation of autoimmune diseases in affected families. *JAMA Intern. Med.* 175 (9), 1518–1526. doi:10.1001/jamainternmed.2015.3528

Labelle-Côté, M., Dusseault, J., Ismaïl, S., Picard-Cloutier, A., Siegel, P. M., and Larose, L. (2011). Nck2 promotes human melanoma cell proliferation, migration and invasion *in vitro* and primary melanoma-derived tumor growth *in vivo*. *BMC Cancer* 11 (1), 443. doi:10.1186/1471-2407-11-443

Langefeld, C. D., Ainsworth, H. C., Cunninghame Graham, D. S., Kelly, J. A., Comeau, M. E., Marion, M. C., et al. (2017). Transancestral mapping and genetic load in systemic lupus erythematosus. *Nat. Commun.* 8, 16021. doi:10.1038/ncomms16021

Lappalainen, T., Sammeth, M., Friedlander, M. R., t Hoen, P. A., Monlong, J., Rivas, M. A., et al. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 501 (7468), 506–511. doi:10.1038/nature12531

Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D. C., Rohland, N., Mallick, S., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. *Nature* 536 (7617), 419–424. doi:10.1038/nature19310

Lee, H. S., Kim, T., Bang, S. Y., Na, Y. J., Kim, I., Kim, K., et al. (2014). Ethnic specificity of lupus-associated loci identified in a genome-wide association study in Korean women. *Ann. Rheum. Dis.* 73 (6), 1240–1245. Epub 2013 Jun 5. doi:10.1136/annrheumdis-2012-202675

Leslie, R., O'Donnell, C. J., and Johnson, A. D. (2014). Grasp: Analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. *Bioinformatics* 30 (12), i185–i194. doi:10.1093/bioinformatics/btu273

Lewis, M. J., and Jawad, A. S. (2017). The effect of ethnicity and genetic ancestry on the epidemiology, clinical features and outcome of systemic lupus erythematosus. *Rheumatol. Oxf.* 56 (1), i67–i77. doi:10.1093/rheumatology/kew399

Li, J., Qiu, D., Chen, Z., Du, W., Liu, J., and Mo, X. (2018). miR-548k regulates CXCL13 expression in myasthenia gravis patients with thymic hyperplasia and in Jurkat cells. *J. Neuroimmunol.* 320, 125–132. doi:10.1016/j.jneuroim.2018.03.021

Li, K., Xu, Y., and Yuan, L. N. (2019a). Down-regulation of miR-5702 is associated with clinical progression and poor prognosis in patients with non-small-cell lung cancer. *Eur. Rev. Med. Pharmacol. Sci.* 23 (5), 2047–2052. doi:10.26355/eurrev_201903_17245

Li, M., Qi, Y., Chen, M., Wang, Z., Zeng, D., Xiao, Y., et al. (2019b). GATA binding protein 3 boosts extracellular ATP hydrolysis and inhibits metastasis of breast cancer by up-regulating Ectonucleoside triphosphate diphosphohydrolase 3. *Int. J. Biol. Sci.* 15 (12), 2522–2537. doi:10.7150/ijbs.35563

Li, Y., Willer, C., Sanna, S., and Abecasis, G. (2009). Genotype imputation. *Annu. Rev. Genomics Hum. Genet.* 10, 387–406. doi:10.1146/annurev.genom.9.081307.164242

Mak, A., and Tay, S. H. (2014). Environmental factors, toxicants and systemic lupus erythematosus. *Int. J. Mol. Sci.* 15 (9), 16043–16056. doi:10.3390/ijms150916043

Marigorta, U. M., and Navarro, A. (2013). High trans-ethnic replicability of GWAS results implies common causal variants. *PLoS Genet.* 9 (6), e1003566. doi:10.1371/journal.pgen.1003566

Molineros, J. E., Yang, W., Zhou, X. J., Sun, C., Okada, Y., Zhang, H., et al. (2017). Confirmation of five novel susceptibility loci for systemic lupus erythematosus (SLE) and integrated network analysis of 82 SLE susceptibility loci. *Hum. Mol. Genet.* 26 (6), 1205–1216. doi:10.1093/hmg/ddx026

Morettin, A., Paris, G., Bouzid, Y., Baldwin, R. M., Falls, T. J., Bell, J. C., et al. (2017). Tudor domain containing protein 3 promotes tumorigenesis and invasive capacity of breast cancer cells. *Sci. Rep.* 7 (1), 5153. doi:10.1038/s41598-017-04955-4

Munz, M., Wohlers, I., Simon, E., Reinberger, T., Busch, H., Schaefer, A. S., et al. (2020). Qtlizer: Comprehensive QTL annotation of GWAS results. *Sci. Rep.* 10 (1), 20417. doi:10.1038/s41598-020-75770-7

Nakagawa, S. (2004). A farewell to Bonferroni: The problems of low statistical power and publication bias. *Behav. Ecol.* 15 (6), 1044–1045. doi:10.1093/beheco/arh107

Ohno, T., Meguro, A., Takeuchi, M., Yamane, T., Teshigawara, T., Kitaichi, N., et al. (2019). Association study of ARMC9 gene variants with vogt-koyanagi-harada disease in Japanese patients. *Ocul. Immunol. Inflamm.* 27 (5), 699–705. doi:10.1080/09273948.2018.1523438

Okumura, T., Ikeda, K., Ujihira, T., Okamoto, K., Horie-Inoue, K., Takeda, S., et al. (2017). Proteasome 26S subunit PSMD1 regulates breast cancer cell growth through p53 protein degradation. *J. Biochem.* 163 (1), 19–29. doi:10.1093/jb/mvx053

Otani, S., Sakurai, T., Yamamoto, K., Fujita, T., Matsuzaki, Y., Goto, Y., et al. (2006). Frequent immune response to a melanocyte specific protein KU-MEL-1 in patients with Vogt-Koyanagi-Harada disease. *Br. J. Ophthalmol.* 90 (6), 773–777. doi:10.1136/bjo.2005.086520

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *BMJ* 316 (7139), 1236–1238. doi:10.1136/bmj.316.7139.1236

Petri, M., Orbai, A. M., Alarcón, G. S., Gordon, C., Merrill, J. T., Fortin, P. R., et al. (2012). Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. *Arthritis Rheum.* 64 (8), 2677–2686. doi:10.1002/art.34473

Piriyapongsa, J., and Jordan, I. K. (2007). A family of human MicroRNA genes from miniature inverted-repeat transposable elements. *PLOS ONE* 2 (2), e203. doi:10.1371/journal.pone.0000203

Pisetsky, D. S., Rovin, B. H., and Lipsky, P. E. (2017). New perspectives in rheumatology: Biomarkers as entry criteria for clinical trials of new therapies for systemic lupus erythematosus: The example of antinuclear antibodies and anti-DNA. *Arthritis Rheumatol.* 69 (3), 487–493. doi:10.1002/art.40008

Popejoy, A. B., and Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature* 538 (7624), 161–164. doi:10.1038/538161a

Ramachandran, S., Ma, T. S., Griffin, J., Ng, N., Foskolou, I. P., Hwang, M. S., et al. (2021). Hypoxia-induced SETX links replication stress with the unfolded protein response. *Nat. Commun.* 12 (1), 3686. doi:10.1038/s41467-021-24066-z

Ramasamy, A., Trabzuni, D., Guelfi, S., Varghese, V., Smith, C., Walker, R., et al. (2014). Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat. Neurosci.* 17 (10), 1418–1428. doi:10.1038/nn.3801

Rung, J., Cauchi, S., Albrechtsen, A., Shen, L., Rocheleau, G., Cavalcanti-Proença, C., et al. (2009). Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat. Genet.* 41 (10), 1110–1115. doi:10.1038/ng.443

Schuenemann, V. J., Peltzer, A., Welte, B., van Pelt, W. P., Molak, M., Wang, C. C., et al. (2017). Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. *Nat. Commun.* 8, 15694. doi:10.1038/ncomms15694

Shi, Y., Qiu, M., Wu, Y., and Hai, L. (2015). MiR-548-3p functions as an anti-oncogenic regulator in breast cancer. *Biomed. Pharmacother.* 75, 111–116. doi:10.1016/j.biopha.2015.07.027

Skol, A. D., Scott, L. J., Abecasis, G. R., and Boehnke, M. (2006). Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* 38 (2), 209–213. doi:10.1038/ng1706

Slattery, M. L., Samowitz, W., Curtin, K., Ma, K. N., Hoffman, M., Caan, B., et al. (2004). Associations among IRS1, IRS2, IGF1, and IGFBP3 genetic polymorphisms and colorectal cancer. *Cancer Epidemiol. Biomarkers Prev.* 13 (7), 1206–1214. doi:10.1158/1055-9965.1206.13.7

Suhre, K., Arnold, M., Bhagwat, A. M., Cotton, R. J., Engelke, R., Raffler, J., et al. (2017). Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.* 8 (1), 14357. doi:10.1038/ncomms14357

Sun, C., Molineros, J. E., Looger, L. L., Zhou, X. J., Kim, K., Okada, Y., et al. (2016). High-density genotyping of immune-related loci identifies new SLE risk variants in individuals with Asian ancestry. *Nat. Genet.* 48, 323–330. doi:10.1038/ng.3496

Tangtanatakul, P., Thumarat, C., Satproedprai, N., Kunhapan, P., Chaiyasung, T., Klinchanhom, S., et al. (2020). Meta-analysis of genome-wide association study identifies FBN2 as a novel locus associated with systemic lupus erythematosus in Thai population. *Arthritis Res. Ther.* 22 (1), 185. doi:10.1186/s13075-020-02276-y

Tishkoff, S. A., Reed, F. A., Friedlaender, F. R., Ehret, C., Ranciaro, A., Froment, A., et al. (2009). The genetic structure and history of Africans and african Americans. *Science* 324(5930), 1035–1044. doi:10.1126/science.1172257

Treamtrakanpon, W., Tantivitayakul, P., Benjachat, T., Somparn, P., Kittikowit, W., Eiam-ong, S., et al. (2012). APRIL, a proliferation-inducing ligand, as a potential marker of lupus nephritis. *Arthritis Res. Ther.* 14 (6), R252. doi:10.1186/ar4095

Tu, Y., Li, F., and Wu, C. (1998). Nck-2, a novel src homology2/3-containing adaptor protein that interacts with the LIM-only protein PINCH and components of growth factor receptor kinase-signaling pathways. *Mol. Biol. Cell* 9 (12), 3367–3382. doi:10.1091/mbc.9.12.3367

Tu, Y., Liang, L., Frank, S. J., and Wu, C. (2001). Src homology 3 domain-dependent interaction of Nck-2 with insulin receptor substrate-1. *Biochem. J.* 354 (2), 315–322. doi:10.1042/0264-6021:3540315

Van De Weghe, J. C., Rusterholz, T. D. S., Latour, B., Grout, M. E., Aldinger, K. A., Shaheen, R., et al. (2017). Mutations in ARMC9, which encodes a basal body protein, cause Joubert syndrome in humans and ciliopathy phenotypes in zebrafish. *Am. J. Hum. Genet.* 101 (1), 23–36. doi:10.1016/j.ajhg.2017.05.010

Voight, B. F., Scott, L. J., Steinthorsdottir, V., Morris, A. P., Dina, C., Welch, R. P., et al. (2010). Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* 42 (7), 579–589. doi:10.1038/ng.609

Wan, L., Zhu, J., and Wu, Q. (2021). Knockdown of DIAPH3 inhibits the proliferation of cervical cancer cells through inactivating mTOR signaling pathway. *J. Oncol.* 2021, 4228241. doi:10.1155/2021/4228241

Wang, J., Zou, Y., Du, B., Li, W., Yu, G., Li, L., et al. (2021a). SNP-mediated lncRNA-ENTPD3-AS1 upregulation suppresses renal cell carcinoma via miR-155/HIF-1α signaling. *Cell Death Dis.* 12 (7), 672. doi:10.1038/s41419-021-03958-4

Wang, M., Futamura, M., Wang, Y., and You, M. (2005). Pas1c1 is a candidate for the mouse pulmonary adenoma susceptibility 1 locus. *Oncogene* 24 (11), 1958–1963. doi:10.1038/sj.onc.1208295

Wang, Y. F., Zhang, Y., Lin, Z., Zhang, H., Wang, T. Y., Cao, Y., et al. (2021b). Identification of 38 novel loci for systemic lupus erythematosus and genetic heterogeneity between ancestral groups. *Nat. Commun.* 12 (1), 772. doi:10.1038/s41467-021-21049-y

Wang, Y. F., Zhang, Y., Zhu, Z., Wang, T. Y., Morris, D. L., Shen, J. J., et al. (2018). Identification of ST3AGL4, MFHAS1, CSNK2A2 and CD226 as loci associated with systemic lupus erythematosus (SLE) and evaluation of SLE genetics in drug repositioning. *Ann. Rheum. Dis.* 77 (7), 1078–1084. doi:10.1136/annrheumdis-2018-213093

Ward, L. D., and Kellis, M. (2012). HaploReg: A resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 40, D930–D934. doi:10.1093/nar/gkr917

Westra, H. J., Peters, M. J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., et al. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* 45 (10), 1238–1243. doi:10.1038/ng.2756

Wohlers, I., Künstner, A., Munz, M., Olbrich, M., Fähnrich, A., Calonga-Solís, V., et al. (2020). An integrated personal and population-based Egyptian genome reference. *Nat. Commun.* 11 (1), 4719. doi:10.1038/s41467-020-17964-1

Wojcik, G. L., Graff, M., Nishimura, K. K., Tao, R., Haessler, J., Gignoux, C. R., et al. (2019). Genetic analyses of diverse populations improves discovery for complex traits. *Nature* 570 (7762), 514–518. doi:10.1038/s41586-019-1310-4

Xia, K., Shabalin, A. A., Huang, S., Madar, V., Zhou, Y.-H., Wang, W., et al. (2011). seeQTL: a searchable database for human eQTLs. *Bioinformatics* 28 (3), 451–452. doi:10.1093/bioinformatics/btr678

Xiao, C., and Rajewsky, K. (2009). MicroRNA control in the immune system: Basic principles. *Cell* 136 (1), 26–36. doi:10.1016/j.cell.2008.12.027

Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). Gcta: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88 (1), 76–82. Epub 2010 Dec 17. doi:10.1016/j.ajhg.2010.11.011

Yelo, E., Bernardo, M. V., Gimeno, L., Alcaraz-García, M. J., Majado, M. J., and Parrado, A. (2008). Dock10, a novel CZH protein selectively induced by interleukin-4 in human B lymphocytes. *Mol. Immunol.* 45 (12), 3411–3418. doi:10.1016/j.molimm.2008.04.003

Yin, X., Kim, K., Suetsugu, H., Bang, S. Y., Wen, L., Koido, M., et al. (2020). Meta-analysis of 208370 East Asians identifies 113 susceptibility loci for systemic lupus erythematosus. *Ann. Rheum. Dis.* 80, 632–640. doi:10.1136/annrheumdis-2020-219209

Yu, C.-H., Pal, L. R., and Moult, J. (2016). Consensus genome-wide expression quantitative trait loci and their relationship with human complex trait disease. *OMICS A J. Integr. Biol.* 20 (7), 400–414. doi:10.1089/omi.2016.0063

Yu, K., Li, Q., Cheng, Q., Huang, C., Zheng, J., Chen, S., et al. (2017). MicroRNA-548j inhibits type I interferon production by targeting ZBTB11 in patients with chronic Hepatitis B. *Biochem. Biophys. Res. Commun.* 488 (4), 628–633. doi:10.1016/j.bbrc.2017.03.109

Yuanfeng, G., Song, J., Li, S., Cui, Y., Lee, C., and Chen, H. (2016). GW27-e0470 mutations in SCAP and AGXT2 revealed by exome sequencing in a pedigree with premature myocardial infarction. *J. Am. Coll. Cardiol.* 68(16_Supplement), C19–C20. doi:10.1016/j.jacc.2016.07.073

Zaitlen, N., Paşaniuc, B., Gur, T., Ziv, E., and Halperin, E. (2010). Leveraging genetic variability across populations for the identification of causal variants. *Am. J. Hum. Genet.* 86 (1), 23–33. doi:10.1016/j.ajhg.2009.11.016

Zhang, C., Xue, Q., Xu, Z., and Lu, C. (2018). MiR-5702 suppresses proliferation and invasion in non-small-cell lung cancer cells via posttranscriptional suppression of ZEB1. *J. Biochem. Mol. Toxicol.* 32 (7), e22163. doi:10.1002/jbt.22163

Zhang, Q., Zhou, X., Wan, M., Zeng, X., Luo, J., Xu, Y., et al. (2021). FoxP3-miR-150-5p/3p suppresses ovarian tumorigenesis via an IGF1R/IRS1 pathway feedback loop. *Cell Death Dis.* 12 (3), 275. doi:10.1038/s41419-021-03554-6

frontiers | Frontiers in Genetics

TYPE Systematic Review
PUBLISHED 19 October 2022
DOI 10.3389/fgene.2022.932764

Check for updates

OPEN ACCESS

EDITED BY
Yunqing Ren,
Zhejiang University, China

REVIEWED BY
Sheng Yang,
Nanjing Medical University, China
Zhongshang Yuan,
Shandong University, China
Guoqing Qian,
Ningbo First Hospital, China

*CORRESPONDENCE
Zhixin Zhang,
zhangzhixin032@163.com
Wenquan Niu,
niuwenquan_shcn@163.com

SPECIALTY SECTION
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

RECEIVED 30 April 2022
ACCEPTED 08 September 2022
PUBLISHED 19 October 2022

CITATION
Yang M, Pang B, Wang Q, Zhang Z and Niu W (2022), The causal association between genetically regulated 25OHD and chronic obstructive pulmonary disease: A meta-analysis and Mendelian randomization study.
Front. Genet. 13:932764.
doi: 10.3389/fgene.2022.932764

COPYRIGHT
© 2022 Yang, Pang, Wang, Zhang and Niu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The causal association between genetically regulated 25OHD and chronic obstructive pulmonary disease: A meta-analysis and Mendelian randomization study

Min Yang[1], Bo Pang[2,3], Qiong Wang[2,3], Zhixin Zhang[4]* and Wenquan Niu[5]*

[1]Department of Acupuncture and Moxibustion, Guang'anmen Hospital, China Academy of Chinese Medical Sciences, Beijing, China, [2]Graduate School, Beijing University of Chinese Medicine, Beijing, China, [3]Department of Pediatrics, China-Japan Friendship Hospital, Beijing, China, [4]International Medical Services, China-Japan Friendship Hospital, Beijing, China, [5]Institute of Clinical Medical Sciences, China-Japan Friendship Hospital, Beijing, China

Backgrounds and objectives: Chronic obstructive pulmonary disease (COPD) is a multifactorial disease under genetic control. We present a meta-analysis to examine the associations of vitamin D binding protein (*VDBP*) gene rs7041 polymorphism with the risk of COPD and changes in circulating 25OHD concentrations.

Methods: A literature search, quality assessment, and data extraction were conducted independently by two investigators. Data are expressed as odds ratio (OR) or weighted mean difference (WMD) with a 95% confidence interval (CI). The inverse variance weighted method (IVW) in R (version 1.1.456) was applied to calculate the Mendelian randomization coefficient.

Results: A total of 13 articles with 3,667 participants were meta-analyzed. The rs7041-GT genotype was associated with a 49% reduced COPD risk (OR: 0.51, 95% CI: 0.30 to 0.88, $p = 0.014$) compared to the rs7041-TT genotype. Carriers of the rs7041-GT genotype had significantly higher concentrations of circulating 25OHD than those with the rs7041-TT genotype (WMD: 0.32 ng/ml, 95% CI: 0.09 to 0.55, $p = 0.006$). Under the assumptions of Mendelian randomization, and assuming a linear logistic relationship between circulating 25OHD and COPD, an inverse association was noted after using *VDBP* gene rs7041 polymorphism as an instrument (WMD: −2.07, 95% CI: −3.72 to −0.41, $p = 0.015$). There was a low probability of publication bias.

Conclusion: We observed significant associations of *VDBP* gene rs7041 polymorphism with the risk of COPD and changes in circulating 25OHD concentrations. Importantly, we found a causal relationship between genetically regulated 25OHD concentrations and COPD risk.

KEYWORDS

chronic obstructive pulmonary disease, vitamin D binding protein, polymorphism, meta-analysis, Mendelian randomization

Introduction

Chronic obstructive pulmonary disease (COPD) places a major burden on individuals and public health systems (Cho et al., 2022). The Global Burden of Disease study estimated that 3.2 million people died from COPD in 2015 worldwide, representing an increase of 11.6% compared with 1990 (GBD 2015 Chronic Respiratory Disease Collaborators, 2017). In China, the overall prevalence of spirometry-defined COPD was 8.6%, accounting for nearly 100 million patients with COPD (Wang et al., 2018). COPD prevention should be an urgent strategy that can be implemented by determining potential risk factors and identifying persons who are at risk of developing COPD and who could be targeted for preventive measures.

COPD is a multifactorial disease under genetic control (Marsh et al., 2006; Agustí et al., 2022). The heritability of COPD has been estimated to be 37.7% (Zhou et al., 2013), which emphasizes a clear rationale for the determination of COPD-susceptibility genes or genetic alterations (Silverman et al., 2011; Silverman, 2020). Dozens of genome-wide association studies have been undertaken to decipher the genetic linings of COPD (Pillai et al., 2009; Hansel et al., 2015; Wain et al., 2017); however, one of the major challenges in investigating genetic determinants is heterogeneity. Currently, the list of candidate genes for COPD is constantly being improved and updated. One of the most widely evaluated genes is the gene coding for the vitamin D binding protein (*VDBP*).

The human *VDBP* gene is localized in chromosome 4q11-q13; it has three commonly recognized haplotypes (GC-1F, GC-1S, and GC-2) and rs7041 polymorphisms are one of the most common non-synonymous single-nucleotide polymorphisms (SNPs) for the haplotypes (Cleve and Constans, 1988). The three haplotypes have been reported to have a diverging affinity to 25OHD, the marker that is the best indicator of circulating vitamin D concentrations (Arnaud and Constans, 1993). In addition, *VDBP* genetic alterations have been found to be associated with vitamin D deficiency (Chishimba et al., 2010; Wood et al., 2011). The susceptibility of *VDBP* haplotypes to COPD risk has been widely evaluated. A meta-analysis by Khanna et al. (2019) showed that the GC-1F haplotype and the GC-1F/1F genotype in the *VDBP* gene impose a significant genetic risk for COPD among Asians. In another meta-analysis by Xie et al. (2015), it was found that the GC-1F homozygote may be a risk-conferring factor for COPD and that the GC-2 homozygote may be a protective factor against COPD. To the best of our knowledge, no pooled evidence exists currently on rs7041 polymorphism in the *VDBP* gene associated with COPD. At present, this association is the subject of much debate. For example, Li et al. (2014) found that homozygous carriers of the rs7041 T allele had enhanced susceptibility to COPD, whereas Ishii et al. (2014) failed to detect any significant link between rs7041 and COPD. The causes of these inconsistent findings could be explored by a comprehensive synthesis of published association studies.

To seek the possible causes and offer research insights, we present a meta-analysis to examine the associations of *VDBP* gene rs7041 polymorphism with the risk of COPD and changes in circulating 25OHD concentrations. In the case of statistical significance being found for both associations, we attempt to explore the possibly causal implications of circulating 25OHD in the development of COPD by adopting the Mendelian randomization technique and using rs7041 polymorphism as an instrument.

Methods

This meta-analysis was conducted followed the guidelines in the preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement (Moher et al., 2009). The PRISMA checklist is provided in Supplementary Table S1.

Search strategy

A literature search was carried out in the three online databases (PubMed, EMBASE, and Web of Science) up March 10, 2022, including articles ahead of publication. The following keywords were used in searching: ("Vitamin D binding protein" OR "VDBP" OR "VDB" OR "DBP" OR "Gc-globulin" OR "Gc-globulin" OR "GC") AND ("COPD" OR "chronic obstructive pulmonary disease") AND ("polymorphism" OR "SNP" OR "variant" OR "variation" OR "mutation" OR "single nucleotide polymorphisms").

Two authors (MY and BP) completed the literature search independently, and any disagreement was solved by discussion. In addition, the references of major reviews or meta-analyses were also checked for potentially eligible articles that were not identified by the two authors.

Inclusion criteria

Articles eligible for inclusion in this meta-analysis must simultaneously meet the following criteria: 1) case-control design; 2) published in the English language; 3) providing genotype counts of *VDBP* gene rs7041 polymorphism in COPD patients and controls or mean or median values of circulating vitamin D concentrations across rs7041 genotypes; 4) a clear definition of COPD [COPD is diagnosed by doctors or according to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) or American Thoracic Society (ATS) guidelines]; and 5) a validated genotype assaying method. Moreover, in cases when more than one article was published using the same or part of the same sample of study participants, the article using the largest sample size was included.

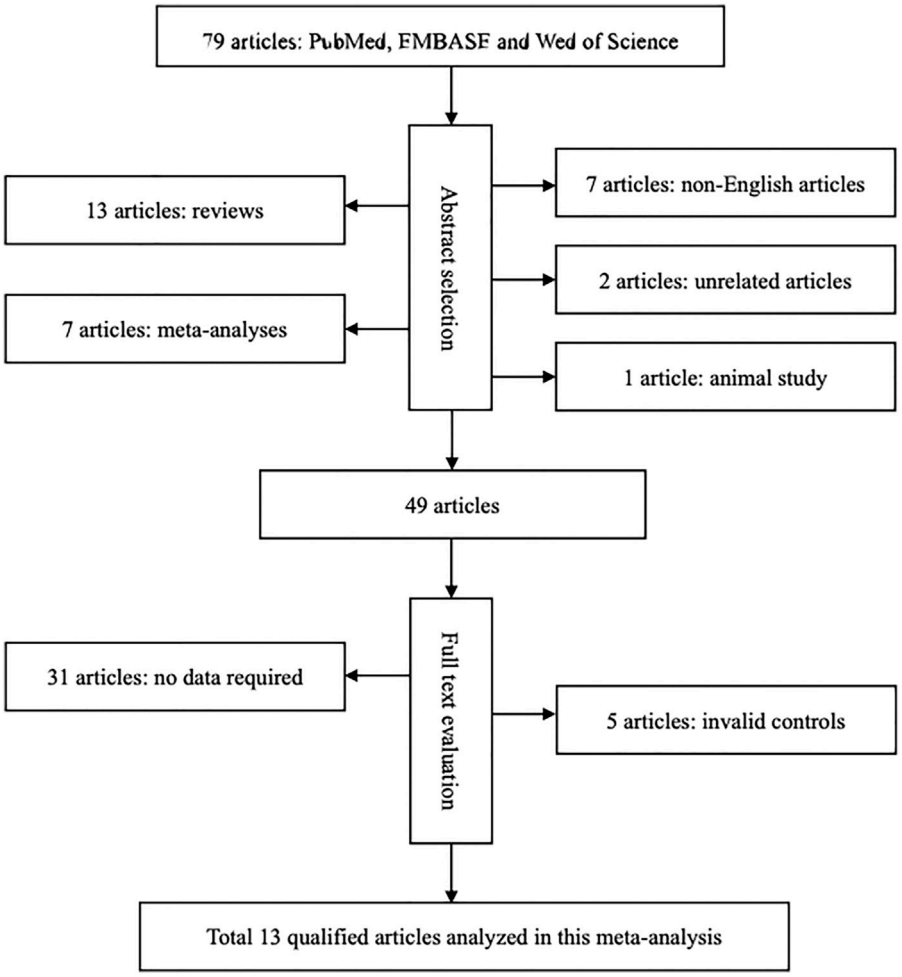


FIGURE 1
Flow diagram for the selection of qualified articles.

Exclusion criteria

Articles were excluded if any one of the following criteria was met: 1) published in form of a review, correspondence, comment, conference abstract, case report, case series, or clinical trial; 2) a clinical outcome other than COPD; and 3) the lack of a control group.

Data extraction

Data were extracted from each eligible article independently by two authors (MY and BP) according to a predefined template, covering the surname of the first author, year of publication, country where the study was performed, ethnicity, sample size, source of controls, matched condition, diagnosis criteria of COPD, genotype counts of rs7041 polymorphism between patients and controls, mean or median 25OHD concentrations across rs7041 genotypes in patients or controls or both, and baseline characteristics of study participants including age, gender, body mass index (BMI), percentage of smokers, smoking exposure (cigarette pack-years), smoking status, forced expiratory volume in 1 s (FEV_1), FEV_1 predicted ratio, forced vital capacity (FVC), FVC predicted ratio, and FEV_1/FVC, if available.

Statistical analyses

STATA software version 14.1 (StataCorp, College Station, TX, United States) was utilized in this meta-analysis. The weighted odds ratio (OR) and the 95% confidence interval (CI) were calculated to quantify the association of *VDBP* gene rs7041 polymorphism with the risk of COPD. The weighted

TABLE 1 Baseline characteristics of all involved studies in this meta-analysis.

First author	Year	Country	Ethnicity	Sample size	Patients	Controls	Match	Diagnosis	COPD stage	Source of controls	NOS score	Age (years)	
												Patients	Controls
Horne	1990	Canada	Caucasian	517	104	413	Yes	Doctor	NA	Population	9	NA	NA
Ishii	2001	Japan	Asian	145	63	82	Yes	ATS	NA	Hospital	8	68.3 (9.9)	NA
Laufs	2004	Iceland	Caucasian	285	102	183	No	ATS	NA	Population	8	71.1	42.9
Ito	2004	Japan	Asian	191	103	88	Yes	ATS	NA	Hospital	8	67.4 (7.8)	60.8 (12.0)
Shen	2010	China	Asian	200	100	100	No	ATS	NA	Hospital	7	62.3 (9.7)	60.9 (8.6)
Janssens	2010	Belgium	Caucasian	414	262	152	Yes	ATS	I-IV	Population	9	66 (60–72)	61 (58–65)
Jung	2014	South Korea	Asian	360	203	157	No	ATS	I-IV	Population	8	67 (9.03)	53 (8.89)
Kukkonen	2014	India	Asian	100	50	50	Yes	Doctor	I-IV	Hospital	8	55.86 (7.42)	55.82 (7.89)
Ishii	2014	Japan	Asian	580	361	219	No	GOLD	I-IV	Hospital	7	68.9 (8.5)	63.1 (12.0)
Al-Azzawi	2017	Egypt	Middle Eastern	160	80	80	Yes	ATS	I-IV	Hospital	8	55.0 (9.0)	53.1 (7.4)
Al-Azzawi (Smokers)	2017	Egypt	Middle Eastern	120	80	40	Yes	ATS	I-IV	Hospital	8	55.0 (9.0)	52.1 (6.5)
Al-Azzawi (Non-smokers)	2017	Egypt	Middle Eastern	120	80	40	Yes	ATS	I-IV	Hospital	8	55.0 (9.0)	55.7 (8.5)
Al-Azzawi (GOLD I)	2017	Egypt	Middle Eastern	96	16	80	Yes	ATS	I	Hospital	8	NA	53.05 (7.4)
Al-Azzawi (GOLD II)	2017	Egypt	Middle Eastern	120	40	80	Yes	ATS	II	Hospital	8	NA	53.05 (7.4)
Al-Azzawi (GOLD III-IV)	2017	Egypt	Middle Eastern	104	24	80	Yes	ATS	III-IV	Hospital	8	NA	53.05 (7.4)
Jolliffe	2018	United Kingdom	Mixed	269	269	0	NA	Doctor	I-IV	NA	7	66.4 (9.5)	NA
Chuaychoo	2018	Thailand	Asian	204	136	68	No	GOLD	NA	Hospital	7	71.0 (8.8)	65.5 (8.9)
Gao	2020	Finland	Caucasian	233	44	189	Yes	GOLD	I-III	Population	9	58.9 (7.1)	53.5 (9.3)
Gao (Non-smokers)	2020	Finland	Caucasian	76	44	32	Yes	GOLD	I-III	Population	9	58.9 (7.1)	54.9 (9.6)
Gao (Control Smokers)	2020	Finland	Caucasian	201	44	157	Yes	GOLD	I-III	Population	9	58.9 (7.1)	52.1 (9.0)
Gao (GOLD I & Smokers)	2020	Finland	Caucasian	210	21	189	Yes	GOLD	I	Population	9	55.1 (8.3)	53.5 (9.3)
Gao (GOLD II-III & Smokers)	2020	Finland	Caucasian	212	23	189	Yes	GOLD	II-III	Population	9	62.7 (5.8)	53.5 (9.3)

Gender (males, %)		BMI (kg/m^2)		Smokers (%)		CPY (pack-years)		FEV_1 (L)		Predicted FEV_1 (%)		FVC (L)		Predicted FVC (%)	
Patients	Controls	Patients	Controls	Patients	Controls	Patients	Controls	Patients	Controls	Patients	Controls	Patients	Controls	Patients	Controls
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
0.95	NA	NA	NA	NA	NA	102.2 (40.4)	NA	NA	NA	NA	NA	NA	NA	NA	NA
0.41	0.57	NA	NA	0.93	NA	38	NA	NA	NA	NA	NA	NA	NA	NA	NA
0.96	0.82	NA	NA	NA	NA	58.3 (29.1)	41.1	1.2 (0.5)	2.7 (0.9)	45.3 (18.4)	88.4 (20.7)	NA	NA	NA	NA
0.72	0.66	NA	NA	NA	NA	123.5 (29.7)	25.6 (13.3)	1.3 (0.3)	3.6 (0.7)	NA	NA	NA	NA	NA	NA
0.82	0.79	NA	NA	1.00	1.00	47.0 (33.0,63.0)	NA	1.8 (0.9)	3.2 (0.8)	61.0 (27.0)	104.0 (15.0)	3.4 (1.0)	4.1 (0.9)	93.0 (22.0)	110.0 (14.0)
0.98	0.94	NA	NA	1.00	1.00	46 (23.4)	30.7 (17.4)	1.4 (0.5)	3.2 (0.6)	53.1 (16.9)	93.9 (12.8)	3.0 (0.8)	4.1 (0.7)	79.6 (19.3)	93.3 (11.9)
0.76	0.76	NA	NA	0.76	NA	23.6 (18.1)	NA	NA	NA	NA	NA	NA	NA	NA	NA
0.92	0.83	NA	NA	1.00	1.00	65.7 (39.3)	51.1 (40.6)	1.7 (0.7)	2.8 (0.8)	60.3 (20.4)	95.8 (15.5)	3.2 (0.9)	3.5 (0.9)	94.6 (18)	99 (15.8)
0.73	0.78	NA	NA	0.50	0.50	46.4 (1.7)	33.7 (1.2)	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	0.50	1.00	46.4 (1.7)	33.7 (1.2)	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	0.50	0.00	46.4 (1.7)	0	NA	NA	NA	NA	NA	NA	NA	NA

(Continued on following page)

10.3389/fgene.2022.932764

TABLE 1 (*Continued*) Baseline characteristics of all involved studies in this meta-analysis.

Gender (males, %)		BMI (kg/m^2)		Smokers (%)		CPY (pack-years)		FEV_1 (L)		Predicted FEV_1 (%)		FVC (L)		Predicted FVC (%)	
Patients	**Controls**	**Patients**	**Controls**	**Patients**	**Controls**	**Patients**	**Controls**	**Patients**	**Controls**	**Patients**	**Controls**	**Patients**	**Controls**	**Patients**	**Controls**
NA	NA	NA	NA	NA	0.50	NA	33.7 (1.2)	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	0.50	NA	33.7 (1.2)	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	0.50	NA	33.7 (1.2)	NA	NA	NA	NA	NA	NA	NA	NA
1	NA	27.7 (6.8)	NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	22.8 (3.9)	24.7 (3.7)	NA	NA	46.5 (54.1)	1.7 (3.9)	1.3 (0.5)	2.6 (0.5)	63.9 (24.6)	111.5 (18.0)	2.5 (0.6)	3.3 (0.6)	92.1 (24.5)	110.9 (21.3)
0.86	0.52	27.3 (3.9)	26.5 (3.7)	1.00	0.83	39.3 (12.9)	28.4 (14.1)	2.7 (0.5)	3.3 (0.7)	75.5 (8.3)	101.5 (13.6)	4.2 (0.8)	3.9 (0.9)	95.8 (5.7)	99.3 (12.2)
0.86	0.38	27.3 (3.9)	26.2 (3.6)	1.00	0.00	39.3 (12.9)	0	2.7 (0.5)	3.2 (0.6)	75.5 (8.3)	106.5 (14.6)	4.2 (0.8)	3.8 (0.8)	95.8 (5.7)	102.3 (12.3)
0.86	0.55	27.3 (3.9)	26.9 (3.8)	1.00	1.00	39.3 (12.9)	28.4 (14.1)	2.7 (0.5)	3.3 (0.8)	75.5 (8.3)	96.5 (12.5)	4.2 (0.8)	4.0 (1.0)	95.8 (5.7)	96.3 (12.1)
0.86	0.52	26.2 (3.3)	26.5 (3.7)	1.00	0.83	35.5 (12.7)	28.4 (14.1)	3.2 (0.5)	3.25 (0.7)	87.3 (5.2)	101.5 (13.6)	4.8 (0.8)	3.9 (0.9)	104.9 (0.1)	99.3 (12.2)
0.87	0.52	28.3 (4.4)	26.5 (3.7)	1.00	0.83	43.1 (13.0)	28.4 (14.1)	2.1 (0.5)	3.25 (0.7)	63.8 (11.4)	101.5 (13.6)	3.6 (0.7)	3.9 (0.9)	86.6 (11.2)	99.3 (12.2)

FEV_1/FVC (%)		rs7041-GG genotype		rs7041-GT genotype		rs7041-TT genotype		25OHD across rs7041 (G/C) genotypes in patients (ng/ml)			25OHD across rs7041 (G/C) genotypes in controls (ng/ml)		
Patients	**Controls**	**Patients**	**Controls**	**Patients**	**Controls**	**Patients**	**Controls**	**rs7041GG**	**rs7041GT**	**rs7041TT**	**rs7041GG**	**rs7041GT**	**rs7041TT**
NA	NA	39	141	0	201	16	72	NA	NA	NA	NA	NA	NA
44.3 (12.7)	NA	20	6	0	32	12	44	NA	NA	NA	NA	NA	NA
NA	NA	38	71	0	86	16	26	NA	NA	NA	NA	NA	NA
45.8 (10.4)	80.1 (7.8)	5	6	0	34	62	47	NA	NA	NA	NA	NA	NA
43.5 (12.8)	92.8 (5.8)	7	7	39	39	54	53	NA	NA	NA	NA	NA	NA
NA	NA	85	54	108	73	60	23	22.1 (7.8)	20.0 (7.0)	16.7 (7.2)	26.3 (8.7)	24.6 (9.1)	20.9 (5.0)
47.5 (10.9)	78.3 (5.1)	9	13	68	65	125	79	NA	NA	NA	NA	NA	NA
NA	NA	8	11	0	25	17	15	NA	NA	NA	NA	NA	NA
51.6 (11.6)	78.6 (5.8)	26	15	131	78	202	125	NA	NA	NA	NA	NA	NA
59 (3.7)	NA	19	25	40	39	22	16	NA	NA	NA	NA	NA	NA
NA	NA	19	10	40	20	22	11	NA	NA	NA	NA	NA	NA
NA	NA	19	16	40	19	22	6	NA	NA	NA	NA	NA	NA
NA	NA	2	25	7	39	8	16	NA	NA	NA	NA	NA	NA
NA	NA	10	25	20	39	11	16	NA	NA	NA	NA	NA	NA
NA	NA	8	25	12	39	4	16	NA	NA	NA	NA	NA	NA
NA	NA	65	NA	147	NA	57	NA	50.3 (28.8)	44.0 (25.5)	40.9 (19.7)	NA	NA	NA
51.2 (12.3)	78.9 (11.4)	13	3	70	28	53	37	NA	NA	NA	NA	NA	NA
63.5 (5.3)	83.1 (5.4)	24	76	17	93	3	20	NA	NA	NA	NA	NA	NA
63.5 (5.3)	84.5 (5.3)	24	8	17	19	3	5	NA	NA	NA	NA	NA	NA
63.5 (5.3)	81.6 (5.4)	24	68	17	74	3	15	NA	NA	NA	NA	NA	NA
67.2 (2.7)	83.1 (5.4)	10	76	10	93	1	20	NA	NA	NA	NA	NA	NA
59.8 (7.8)	83.1 (5.4)	14	76	7	93	2	20	NA	NA	NA	NA	NA	NA

Abbreviations: COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in one second; FVC, forced vital capacity; GOLD, Global Initiative for Chronic Obstructive Lung Disease; ATS, American Thoracic Society; CPY, cigarette pack-years; NOS, Newcastle–Ottawa scale; NA, not available.

10.3389/fgene.2022.932764

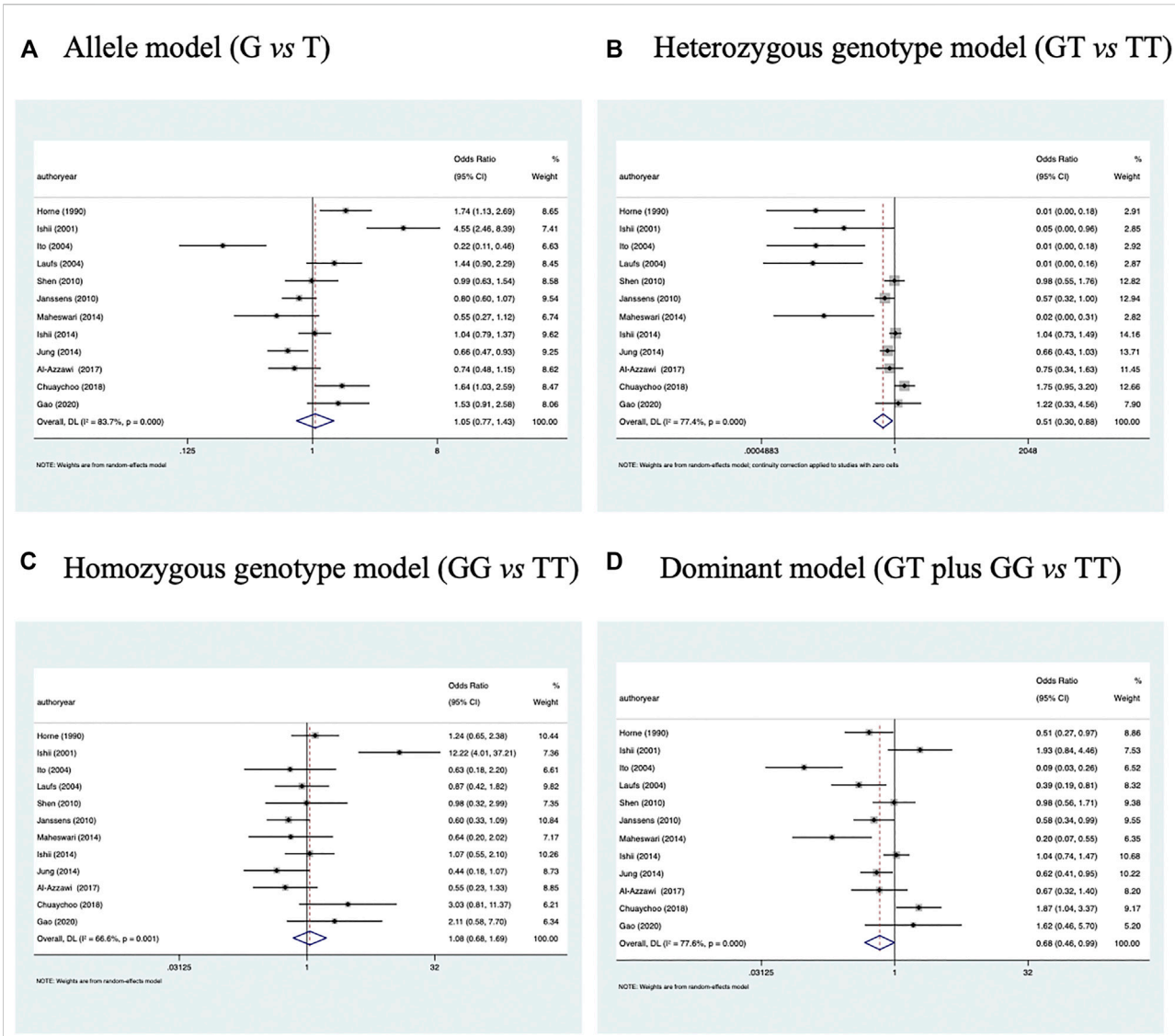


FIGURE 2
Forest plots of *VDBP* gene rs7041 polymorphism with COPD under four models of inheritance. **(A)** Allele model (G vs. T), **(B)** Heterozygous genotype model (GT vs. TT), **(C)** Homozygous genotype model (GG vs. TT), **(D)** Dominant model (GT plus GG vs. TT).

mean difference (WMD) and 95% CI were calculated to quantify the changes of circulating 25OHD concentrations between genotypes of this polymorphism in COPD patients or controls or both ($p < 0.05$; statistically significant). The Hardy–Weinberg equilibrium for *VDBP* gene rs7041 polymorphism was calculated using the R package "GWASExactHW" (version 1.1.456), and $p > 0.05$ indicated no deviation from the Hardy–Weinberg equilibrium.

The fraction of variation owing to heterogeneity was estimated using *I*-squared (I^2). A larger I^2 value is indicative of a higher probability of between-study heterogeneity. Significance for I^2 was set at 50% (Higgins et al., 2003). Whatever the magnitude of heterogeneity, the random-effects model was used to calculate effect–size estimates. The causes for between-study heterogeneity were explored from clinical and methodological aspects by subgroup analyses and meta-regression analyses.

Cumulative analyses were performed to evaluate the influence of the first publication on the association between *VDBP* gene rs7041 polymorphism and the risk of COPD on subsequent publications on the same subject over time. In addition, sensitivity analyses were performed to evaluate the contribution of single publications to pooled effect-size estimates by sequentially omitting one publication each time and deriving estimates from the remaining publications.

Publication bias was evaluated using Begg's funnel plots and Egger's regression asymmetry tests. The Egger test can detect funnel plot asymmetry by quantifying the probability of publication bias at a significance level of 10%. In addition, the

TABLE 2 Subgroup analyses of *VDBP* gene rs7041 polymorphism with COPD risk under 4 models of inheritance.

Subgroup	Allele model (G vs. T)					Heterozygous model (GT vs. TT)				Homozygous model (GG vs. TT)				Dominant model (GG plus GT vs. TT)			
	Studies	OR	95% CI	*p*	I^2	OR	95% CI	*p*	I^2	OR	95% CI	*p*	I^2	OR	95% CI	*p*	I^2
By sample size																	
Total sample size <220	6	0.95	0.48 to 1.90	0.893	89.7%	0.35	0.14 to 0.91	0.072	80.1%	1.40	0.52 to 3.79	0.512	79.0%	0.63	0.27 to 1.46	0.283	86.6%
Total sample size ≥220	6	1.09	0.81 to 1.47	0.552	73.8%	0.53	0.27 to 1.03	0.062	78.0%	0.87	0.61 to 1.26	0.463	29.3%	0.68	0.48 to 0.95	0.025	54.8%
By race																	
Asian	7	0.96	0.57 to 1.59	0.861	88.5%	0.64	0.34 to 1.2	0.166	78.1%	1.29	0.57 to 2.89	0.544	76.8%	0.70	0.39 to 1.26	0.233	85.1%
Caucasian	4	1.29	0.86 to 1.94	0.224	74.0%	0.16	0.03 to 0.97	0.046	81.8%	0.94	0.60 to 1.48	0.788	31.6%	0.56	0.38 to 0.83	0.004	19.9%
Middle Eastern	1	0.74	0.48 to 1.15	0.180	NA	0.75	0.34 to 1.63	0.462	NA	0.55	0.23 to 1.33	0.186	NA	0.67	0.32 to 1.40	0.286	NA
By match																	
No	5	1.07	0.79 to 1.45	0.657	67.5%	0.88	0.51 to 1.52	0.644	76.3%	0.94	0.58 to 1.54	0.808	33.2%	0.88	0.57 to 1.35	0.552	72.9%
Yes	7	1.00	0.57 to 1.76	0.992	89.1%	0.19	0.06 to 0.58	0.004	75.6%	1.18	0.57 to 2.46	0.651	77.6%	0.53	0.28 to 1.02	0.058	78.1%
Diagnosis																	
ATS	7	0.93	0.58 to 1.49	0.762	87.8%	0.44	0.23 to 0.85	0.014	72.4%	0.95	0.47 to 1.94	0.894	77.5%	0.59	0.36 to 0.97	0.036	75.8%
Doctor	2	1.02	0.33 to 3.13	0.980	86.4%	0.01	0.00 to 0.10	<0.001	0.0%	1.06	0.60 to 1.86	0.839	0.0%	0.35	0.14 to 0.88	0.025	58.6%
GOLD	3	1.30	0.95 to 1.78	0.102	44.7%	1.20	0.88 to 1.64	0.260	3.8%	1.50	0.82 to 2.76	0.190	12.0%	1.32	0.87 to 2.01	0.195	33.5%
Source of controls																	
Hospital	8	0.95	0.63 to 1.45	0.819	86.2%	0.64	0.36 to 1.14	0.132	75.3%	1.17	0.60 to 2.28	0.640	74.8%	0.69	0.41 to 1.19	0.182	82.8%
Population	4	1.24	0.76 to 2.03	0.392	80.8%	0.17	0.03 to 1.02	0.052	82.6%	0.94	0.55 to 1.63	0.835	40.6%	0.58	0.40 to 0.84	0.004	23.6%

Abbreviations: COPD, chronic obstructive pulmonary disease; OR, odds ratio; 95% CI, 95% confidence interval; GOLD, Global Initiative for Chronic Obstructive Lung Disease; ATS, American Thoracic Society; I^2, inconsistency index; NA, not available.

10.3389/fgene.2022.932764

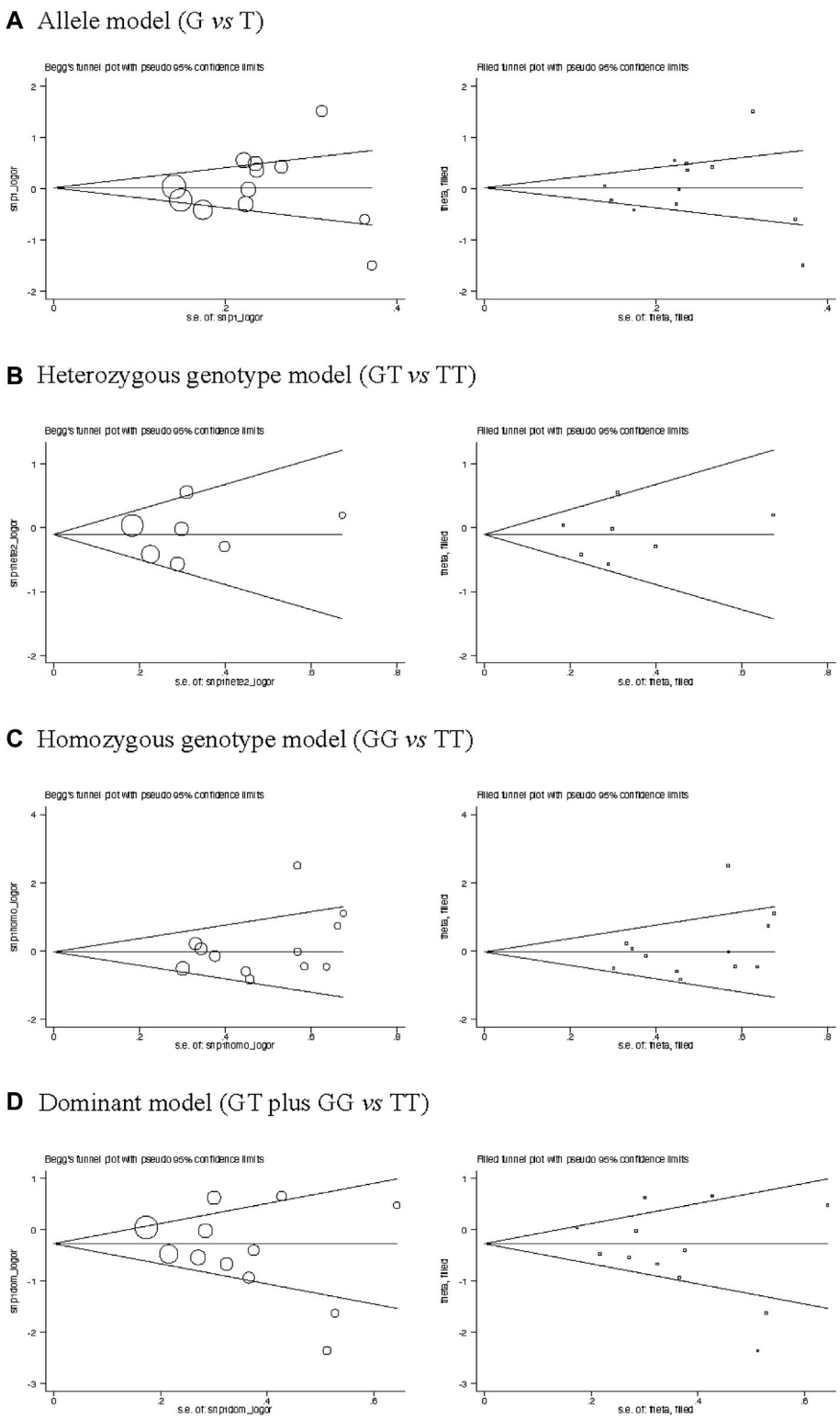


FIGURE 3
Funnel plots of *VDBP* gene rs7041 polymorphism with COPD under four models of inheritance. **(A)** Allele model (G vs. T), **(B)** Heterozygous genotype model (GT vs. TT), **(C)** Homozygous genotype model (GG vs. TT), **(D)** Dominant model (GT plus GG vs. TT).

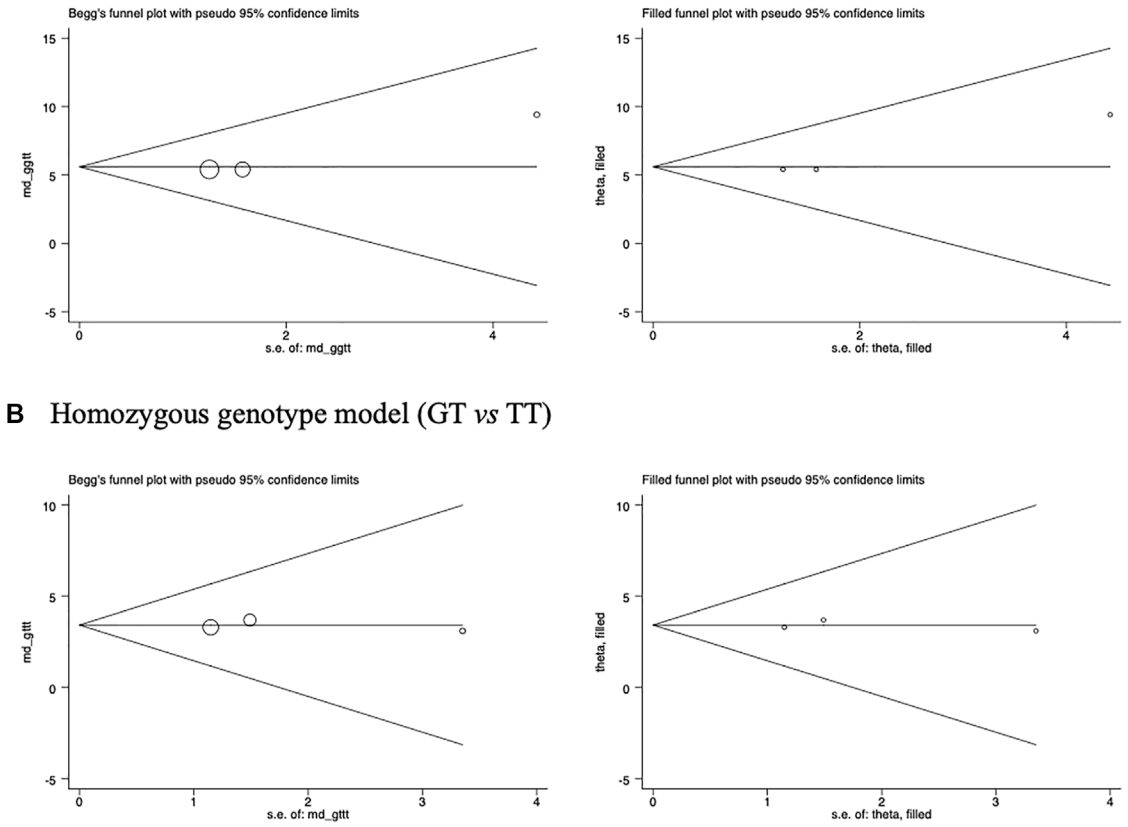


FIGURE 4
Funnel plots of the changes in 25OHD concentrations across rs7041 genotypes. **(A)** Homozygous genotype model (GG vs. TT), **(B)** Heterozygous genotype model (GT vs. TT).

trim-and-fill method was used to estimate the number of potentially missing publications leading to publication bias and to derive corrected pooled estimates.

The Mendelian randomization technique was employed to estimate the causal influencer of a modifiable risk factor from observational data on clinical outcomes (Smith and Ebrahim, 2003). The inverse variance weighted method (IVW) (Burgess et al., 2013) in the R programming environment (version 1.1.456) was applied to calculate the Mendelian randomization coefficient in this meta-analysis.

Results

Eligible studies

The initial search yielded 79 potentially relevant publications. Application of the inclusion and exclusion criteria revealed 13 eligible articles involving 3,389 participants in the final analysis (Horne et al., 1990; Ishii et al., 2001; Ito et al., 2004; Laufs et al., 2004; Janssens et al., 2010; Shen et al., 2010; Ishii et al., 2014; Jung et al., 2014; Maheswari et al., 2014; Al-Azzawi et al., 2017; Chuaychoo et al., 2018; Jolliffe et al., 2018; Gao et al., 2020). The selection process with specific reasons for exclusion is presented in Figure 1. Finally, data from the 3,667 participants were pooled. Data on circulating 25OHD changes across the rs7041 genotypes were extracted from two articles involving 540 patients and 152 controls (Janssens et al., 2010; Jolliffe et al., 2018).

Study characteristics

The baseline characteristics of the study populations are summarized in Table 1. Of the 13 eligible articles, seven were performed among Asians (Ishii et al., 2001; Ito et al., 2004; Shen

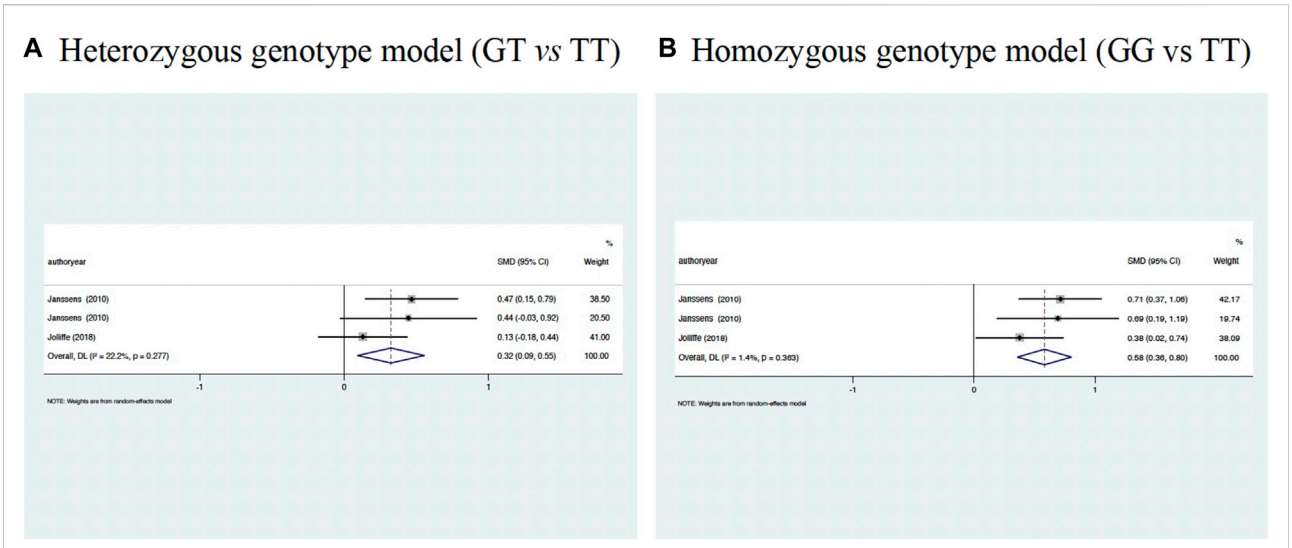


FIGURE 5
Circulating 25OHD changes across *VDBP* gene rs7041 genotypes. **(A)** Heterozygous genotype model (GT vs. TT), **(B)** Homozygous genotype model (GG vs. TT).

et al., 2010; Ishii et al., 2014; Jung et al., 2014; Maheswari et al., 2014; Chuaychoo et al., 2018), four among Caucasians (Horne et al., 1990; Laufs et al., 2004; Janssens et al., 2010; Gao et al., 2020), one among a Middle Eastern population (Maheswari et al., 2014), and one among a mixed population (Jolliffe et al., 2018). Total sample sizes ranged from 100 to 517. A total of seven articles were conducted with matched patients and controls. There was no deviation from the Hardy–Weinberg equilibrium for *VDBP* gene rs7041 polymorphism. Table 1 also shows the quality assessment of all the qualified articles by using the Newcastle–Ottawa scale (NOS) tool for cohort studies.

Overall analysis: *VDBP* and COPD

As shown in Figure 2, the association of *VDBP* gene rs7041 polymorphism with the risk of COPD was tested under four genetic models: allele model (rs7041-G vs. rs7041-T); homozygous genotype model (rs7041-GG vs. rs7041-TT); heterozygous genotype model (rs7041-GT vs. rs7041-TT); and dominant model (rs7041-GG + GT vs. rs7041-TT). This association was significant under the heterozygous genotype model and the dominant model, with the odds of having COPD being 0.51 (95% CI: 0.30 to 0.88, p = 0.014) and 0.68 (95% CI: 0.46 to 0.99, p = 0.047), respectively; however, there was moderate evidence of heterogeneity between studies for both models, with a corresponding I^2 value of 77.4% and 77.6%, respectively.

Subgroup analyses: *VDBP* and COPD

Given the significant heterogeneity in the overall analyses, explorations were undertaken by subgroup analyses according to ethnicity, sample size, source of controls, matched condition, and diagnosis criteria of COPD, respectively (Table 2). Under the heterozygous genotype model, the association of *VDBP* gene rs7041 polymorphism with COPD was potentiated in Caucasians (OR: 0.16, 95% CI: 0.03 to 0.97, p = 0.046), in studies with matched patients and controls (OR: 0.19, 95% CI: 0.06 to 0.58, p = 0.004), in studies with doctor-diagnosed COPD (OR: 0.01, 95% CI: 0.00 to 0.10, p < 0.001), and adopting ATS criteria (OR: 0.44, 95% CI: 0.23 to 0.85, p = 0.014). Under the dominant model, COPD risk was significantly reduced in large studies (OR: 0.68, 95% CI: 0.48 to 0.95, p = 0.025), in studies enrolling population-based controls (OR: 0.58, 95% CI: 0.40 to 0.84, p = 0.004), in studies involving Caucasian populations (OR: 0.56, 95% CI: 0.38 to 0.83, p = 0.004), in studies with doctor-diagnosed COPD (OR: 0.35, 95% CI: 0.14 to 0.88, p = 0.025), and in ATS-criteria-based COPD (OR: 0.59, 95% CI: 0.36 to 0.97, p = 0.036). As shown in Table 2, the heterogeneity of each subgroup of race was still significant with a corresponding I^2 value of 78.1% in the Asian group and 81.8% in the Caucasian group, indicating that race was not the cause of heterogeneity under the heterozygous genotype model. Under the allele model and the homozygous genotype model, there was no noticeable significance across all subgroups.

Meta-regression analyses: *VDBP* and COPD

Further explorations on the causes of heterogeneity were performed using meta-regression analyses via modeling age, male composition, BMI, percentage of smokers, smoking exposure (cigarette pack-years), FEV_1, FEV_1 predicted ratio, FVC, FVC predicted ratio, and FEV_1/FVC in cases and controls under the four genetic models. However, there was no detectable significance for all factors ($p > 0.05$).

Cumulative and sensitivity analyses: *VDBP* and COPD

The impact of the first published article on subsequent articles was not significant for the association between rs7041 polymorphism and COPD in cumulative analyses (Supplementary Figure S1). In addition, the impact of any single article on pooled estimates was also not significant in sensitivity analyses (Supplementary Figure S2).

Publication bias: *VDBP* and COPD

The Begg's funnel plots seemed symmetrical for the association of rs7041 polymorphism with COPD risk under the allele, homozygous genotype, heterozygous genotype, and dominant models (Figure 3). This symmetry was confirmed by the Egger tests, with a corresponding probability of 0.725, 0.869, 0.249, and 0.276, respectively, indicating a low likelihood of publication bias.

25OHD changes and rs7041 genotypes

Changes in circulating 25OHD concentrations across rs7041 genotypes were assessed. As shown in Figure 3, circulating 25OHD concentrations were significantly elevated in rs7041-GT genotype carriers (WMD: 0.32 ng/ml, 95% CI: 0.09 to 0.55, $p = 0.006$) and rs7041-GG genotype carriers (WMD: 0.58 ng/ml, 95% CI: 0.30 to 0.80, $p < 0.001$) compared to rs7041-TT genotype carriers, with low evidence of heterogeneity ($I^2 = 22.2$ and 1.4%, respectively), indicating no significant heterogeneity. In other words, race was not a significant cause of heterogeneity in our meta-analysis. The Begg's funnel plots seemed symmetrical for the association of rs7041 polymorphism with the changes of 25OHD concentrations under homozygous genotype and heterozygous genotype models (Figure 4).

Mendelian randomization analyses

Analogous to a randomized controlled trial, Mendelian randomization is developed as a viable strategy to obtain unconfounded and unbiased estimates of causal relevance from observational data. To infer a causal relationship between exposure and disease outcome, two important prerequisites must be met: 1) the selected genes are significantly correlated with intermediate phenotypes or exposure; and 2) the selected genes are significantly correlated with the condition of the disease outcome (Sheehan et al., 2008). To apply R package's "IVW" for Mendelian randomization analyses, we need the data consisting of the OR and 95% CI of rs7041 polymorphism across COPD patients and controls as well as the OR and 95% CI of circulating 25OHD changes across rs7041 polymorphism under the background of case control studies about COPD patients and controls as shown in Figure 2 and Figure 5, which show that the two important prerequisites mentioned above were both met properly. Specifically, the causal estimate was obtained by regression of the associations with the outcome on the associations with the risk factor, with the intercept set to zero and weights being the inverse variances of the associations with the outcome when the "IVW" package was applied (Burgess et al., 2013). With a single genetic variant, this was simply the ratio method (Burgess et al., 2013). Therefore, assuming a linear logistic relationship between circulating 25OHD concentrations and the risk of COPD, an inverse association (WMD: −2.07, 95% CI: −3.72 to −0.41, $p = 0.015$) was noted after using *VDBP* gene rs7041 polymorphism as an instrument for Mendelian randomization analyses.

Discussion

In this meta-analysis, we aimed to evaluate the association of *VDBP* gene rs7041 polymorphism with the risk of COPD and circulating 25OHD changes. After integrating the results of 13 articles and 3,389 participants, we observed significant associations of this polymorphism with the risk of COPD and changes in circulating 25OHD concentrations. Importantly, under the rationale of Mendelian randomization, we found a causal relationship between genetically regulated 25OHD concentrations and COPD risk. To the best of our knowledge, this is the first meta-analysis to date that has interrogated the causality between circulating 25OHD and COPD in the literature.

The association of *VDBP* gene rs7041 polymorphism with COPD risk has been widely studied. Li et al. (2014) found that COPD patients were at high risk of vitamin D deficiency, and that carrying the rs7041-T allele had an impact on serum 25OHD concentrations that were closely related to COPD susceptibility. In contrast, Ishii et al. (2014) reported that distributions of rs7041 genotypes were comparable between COPD patients

and controls. This inconsistency is likely due to inadequate sample size, patient selection, or lack of adjustment for confounders. To address this point, we performed subgroup analyses and found that the association between the rs7041-GT genotype and COPD risk was significant in populations of Caucasian origin and in studies with matched patients and controls. In view of this ethnic difference, we suggest building a candidate list of susceptible genes for COPD in each ethnic group.

Growing evidence shows that vitamin D is a promising biomarker in the development of COPD (Fu et al., 2021). The results of Fu et al. (2021) showed that vitamin D was inversely correlated with inflammatory signaling in patients with COPD, and that vitamin D may be a vital mediator of the progress of COPD in patients with low vitamin D levels. *VDBP* gene rs7041 polymorphism is responsible for the binding and transport of vitamin D analogs, as well as certain immune functions (Rozmus et al., 2020; Abdelaziz et al., 2021; Rozmus et al., 2022). In agreement with the findings of this meta-analysis, many studies have supported the close relationship between *VDBP* genetic alterations and circulating 25OHD concentrations. For instance, Alharazy et al. (2021) showed that the mutation of rs7041 polymorphism was associated with total 25OHD concentrations in postmenopausal women in Saudi Arabia, and the *post hoc* test indicated that total 25OHD concentrations were lower in carriers of the rs7041-TT genotype than carriers of the rs7041-GG genotype. In a separate study, Al-Daghri et al. (2019) reported that median 25OHD concentrations in carriers of the rs7041-GG genotype were significantly higher than in rs7041-TT genotype counterparts. In the present meta-analysis, we found that circulating 25OHD concentrations were significantly higher in carriers of the rs7041-GT genotype than those with rs7041-TT genotype and, importantly, that genetically regulated 25OHD in circulation played a causal role in the pathogenesis of COPD by means of the Mendelian randomization technique. Nevertheless, we concede that our findings are preliminary and that further independent validations are necessary.

Despite the clear strengths of this meta-analysis, including the large sample size and the adoption of the Mendelian randomization technique to infer causality between circulating 25OHD and COPD, several limitations should be acknowledged. First, caution should be taken when interpreting the relationship of circulating 25OHD concentrations with COPD because only two studies, with only 440 COPD patients and 152 controls, were available for analysis. Second, only articles published in the English language were retrieved, which might yield possible selection bias, even though our funnel plots and statistical tests revealed that this bias was unlikely. Third, although several subgroup analyses were conducted to explore potential heterogeneity, including analyses grouped by race, these subgroup analyses could not demonstrate that the target loci are conserved across different ancestries, and we only chose *VDBP* gene rs7041 polymorphism as an instrument for Mendelian randomization analysis, while other *VDBP* gene polymorphisms were not involved, thus limiting the interpretation of the pooled estimates.

In summary, we observed significant associations of *VDBP* gene rs7041 polymorphism with the risk of COPD and changes in circulating 25OHD concentrations. Importantly, under the rationale of Mendelian randomization, we found a causal relationship between genetically regulated 25OHD concentrations and COPD risk. Furthermore, our findings have highlighted the importance of employing the Mendelian randomization technique to pinpoint causal biomarkers and have provided clues regarding the causative pathways responsible for the biological regulation of 25OHD concentrations in COPD, thereby shedding new light in the physiology of this disease.

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

WN and ZZ planned and designed the study; WN and MY searched the literature, selected the articles, and abstracted the data; BP and QW performed data preparation and quality control; MY and WN analyzed the data and wrote the manuscript.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.932764/full#supplementary-material

References

Abdelaziz, M. H., Ji, X., Wan, J., Abouelnazar, F. A., Abdelwahab, S. F., and Xu, H. (2021). Mycobacterium-induced Th1, helminths-induced Th2 cells and the potential vaccine candidates for allergic asthma: Imitation of natural infection. *Front. Immunol.* 12, 696734. doi:10.3389/fimmu.2021.696734

Agustí, A., Melén, E., DeMeo, D. L., Breyer-Kohansal, R., and Faner, R. (2022). Pathogenesis of chronic obstructive pulmonary disease: Understanding the contributions of gene-environment interactions across the lifespan. *Lancet. Respir. Med.* 10, 512–524. doi:10.1016/S2213-2600(21)00555-5

Al-Azzawi, M. A., Ghoneim, A. H., and Elmadbouh, I. (2017). Evaluation of vitamin D, vitamin D binding protein gene polymorphism with oxidant - antioxidant profiles in chronic obstructive pulmonary disease. *J. Med. Biochem.* 36 (4), 331–340. doi:10.1515/jomb-2017-0012

Al-Daghri, N. M., Mohammed, A. K., Bukhari, I., Rikli, M., Abdi, S., Ansari, M. G. A., et al. (2019). Efficacy of vitamin D supplementation according to vitamin D-binding protein polymorphisms. *Nutrition* 63-64, 148–154. doi:10.1016/j.nut.2019.02.003

Alharazy, S., Naseer, M. I., Alissa, E., Robertson, M. D., Lanham-New, S., Alqahtani, M. H., et al. (2021). Association of SNPs in GC and CYP2R1 with total and directly measured free 25-hydroxyvitamin D in multi-ethnic postmenopausal women in Saudi Arabia. *Saudi J. Biol. Sci.* 28 (8), 4626–4632. doi:10.1016/j.sjbs.2021.04.071

Arnaud, J., and Constans, J. (1993). Affinity differences for vitamin D metabolites associated with the genetic isoforms of the human serum carrier protein (DBP). *Hum. Genet.* 92 (2), 183–188. doi:10.1007/BF00219689

Burgess, S., Butterworth, A., and Thompson, S. G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* 37 (7), 658–665. doi:10.1002/gepi.21758

Chishimba, L., Thickett, D. R., Stockley, R. A., and Wood, A. M. (2010). The vitamin D axis in the lung: A key role for vitamin D-binding protein. *Thorax* 65 (5), 456–462. doi:10.1136/thx.2009.128793

Cho, M. H., Hobbs, B. D., and Silverman, E. K. (2022). Genetics of chronic obstructive pulmonary disease: Understanding the pathobiology and heterogeneity of a complex disorder. *Lancet. Respir. Med.* 10, 485–496. doi:10.1016/S2213-2600(21)00510-5

Chuaychoo, B., Tungtrongchitr, R., Kriengsinyos, W., Tuntipopipat, S., On-Nom, N., and Chupeerach, C. (2018). Correlation of vitamin D binding protein gene polymorphism and protein levels in chronic obstructive pulmonary disease compared with non-chronic obstructive pulmonary disease subjects. *Per. Med.* 15 (5), 371–379. doi:10.2217/pme-2018-0005

Cleve, H., and Constans, J. (1988). The mutants of the vitamin-D-binding protein: More than 120 variants of the GC/DBP system. *Vox Sang.* 54 (4), 215–225. doi:10.1111/j.1423-0410.1988.tb03908.x

Fu, L., Fei, J., Tan, Z. X., Chen, Y. H., Hu, B., Xiang, H. X., et al. (2021). Low vitamin D status is associated with inflammation in patients with chronic obstructive pulmonary disease. *J. Immunol.* 206 (3), 515–523. doi:10.4049/jimmunol.2000964

Gao, J., Törölä, T., Li, C. X., Ohlmeier, S., Toljamo, T., Nieminen, P., et al. (2020). Sputum vitamin D binding protein (VDBP) GC1S/1S genotype predicts airway obstruction: A prospective study in smokers with COPD. *Int. J. Chron. Obstruct. Pulmon. Dis.* 15, 1049–1059. doi:10.2147/COPD.S234464

GBD 2015 Chronic Respiratory Disease Collaborators (2017). Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: A systematic analysis for the global burden of disease study 2015. *Lancet. Respir. Med.* 5 (9), 691–706. doi:10.1016/S2213-2600(17)30293-X

Hansel, N. N., Pare, P. D., Rafaels, N., Sin, D. D., Sandford, A., Daley, D., et al. (2015). Genome-wide association study identification of novel loci associated with airway responsiveness in chronic obstructive pulmonary disease. *Am. J. Respir. Cell Mol. Biol.* 53 (2), 226–234. doi:10.1165/rcmb.2014-0198OC

Higgins, J. P., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ Clin. Res. ed)* 327 (7414), 557–560. doi:10.1136/bmj.327.7414.557

Horne, S. L., Cockcroft, D. W., and Dosman, J. A. (1990). Possible protective effect against chronic obstructive airways disease by the GC2 allele. *Hum. Hered.* 40 (3), 173–176. doi:10.1159/000153926

Ishii, T., Keicho, N., Teramoto, S., AzumA, A., Kudoh, S., Fukuchi, Y., et al. (2001). Association of Gc-globulin variation with susceptibility to COPD and diffuse panbronchiolitis. *Eur. Respir. J.* 18 (5), 753–757. doi:10.1183/09031936.01.00094401

Ishii, T., Motegi, T., Kamio, K., Gemma, A., and Kida, K. (2014). Association of group component genetic variations in COPD and COPD exacerbation in a Japanese population. *Respirology* 19 (4), 590–595. doi:10.1111/resp.12277

Ito, I., Nagai, S., Hoshino, Y., Muro, S., Hirai, T., Tsukino, M., et al. (2004). Risk and severity of COPD is associated with the group-specific component of serum globulin 1F allele. *Chest* 125 (1), 63–70. doi:10.1378/chest.125.1.63

Janssens, W., Bouillon, R., Claes, B., Carremans, C., Lehouck, A., Buysschaert, I., et al. (2010). Vitamin D deficiency is highly prevalent in COPD and correlates with variants in the vitamin D-binding gene. *Thorax* 65 (3), 215–220. doi:10.1136/thx.2009.120659

Jolliffe, D. A., James, W. Y., Hooper, R. L., Barnes, N. C., Greiller, C. L., Islam, K., et al. (2018). Prevalence, determinants and clinical correlates of vitamin D deficiency in patients with chronic obstructive pulmonary disease in London, UK. *J. Steroid Biochem. Mol. Biol.* 175, 138–145. doi:10.1016/j.jsbmb.2017.01.019

Jung, J. Y., Choi, D. P., Won, S., Lee, Y., Shin, J. H., Kim, Y. S., et al. (2014). Relationship of vitamin D binding protein polymorphisms and lung function in Korean chronic obstructive pulmonary disease. *Yonsei Med. J.* 55 (5), 1318–1325. doi:10.3349/ymj.2014.55.5.1318

Khanna, R., Nandy, D., and Senapati, S. (2019). Systematic review and meta-analysis to establish the association of common genetic variations in vitamin D binding protein with chronic obstructive pulmonary disease. *Front. Genet.* 10, 413. doi:10.3389/fgene.2019.00413

Laufs, J., Andrason, H., Sigvaldason, A., Halapi, E., Thorsteinsson, L., Jonasson, K., et al. (2004). Association of vitamin D binding protein variants with chronic mucus hypersecretion in Iceland. *Am. J. Pharmacogenomics* 4 (1), 63–68. doi:10.2165/00129785-200404010-00007

Li, X., Liu, X., Xu, Y., Xiong, W., Zhao, J., Ni, W., et al. (2014). The correlation of vitamin D level and vitamin D-binding protein gene polymorphism in chronic obstructive pulmonary disease. *Zhonghua Nei Ke Za Zhi* 53 (4), 303–307.

Maheswari, K., Choudhary, M., and Javid, S. (2014). Association of Vitamin D Binding protein gene polymorphism with serum 25-hydroxy vitamin D levels in COPD. *Online Int. Interdiscipl Res. J.* 4, 46–55.

Marsh, S., Aldington, S., Shirtcliffe, P., Weatherall, M., and Beasley, R. (2006). Smoking and COPD: What really are the risks? *Eur. Respir. J.* 28 (4), 883–884. doi:10.1183/09031936.06.00074806

Moher, D., Liberati, A., Tetzlaff, J., and Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ* 339, b2535. doi:10.1136/bmj.b2535

Pillai, S. G., Ge, D., Zhu, G., Kong, X., Shianna, K. V., Need, A. C., et al. (2009). A genome-wide association study in chronic obstructive pulmonary disease (COPD): Identification of two major susceptibility loci. *PLoS Genet.* 5 (3), e1000421. doi:10.1371/journal.pgen.1000421

Rozmus, D., Ciesielska, A., Płomiński, J., Grzybowski, R., Fiedorowicz, E., Kordulewska, N., et al. (2020). Vitamin D binding protein (VDBP) and its gene polymorphisms-the risk of malignant tumors and other diseases. *Int. J. Mol. Sci.* 21 (21), e7822. doi:10.3390/ijms21217822

Rozmus, D., Płomiński, J., Augustyn, K., and Cieślińska, A. (2022). rs7041 and rs4588 polymorphisms in vitamin D binding protein gene (VDBP) and the risk of diseases. *Int. J. Mol. Sci.* 23 (2), 933. doi:10.3390/ijms23020933

Sheehan, N. A., Didelez, V., Burton, P. R., and Tobin, M. D. (2008). Mendelian randomisation and causal inference in observational epidemiology. *PLoS Med.* 5 (8), e177. doi:10.1371/journal.pmed.0050177

Shen, L. H., Zhang, X. M., Su, D. J., Yao, S. P., Yu, B. Q., Wang, H. W., et al. (2010). Association of vitamin D binding protein variants with susceptibility to chronic obstructive pulmonary disease. *J. Int. Med. Res.* 38 (3), 1093–1098. doi:10.1177/147323001003800337

Silverman, E. K. (2020). Genetics of COPD. *Annu. Rev. Physiol.* 82, 413–431. doi:10.1146/annurev-physiol-021317-121224

Silverman, E. K., Vestbo, J., Agusti, A., Anderson, W., Bakke, P. S., Barnes, K. C., et al. (2011). Opportunities and challenges in the genetics of COPD 2010: An international COPD genetics conference report. *COPD* 8 (2), 121–135. doi:10.3109/15412555.2011.558864

Smith, G. D., and Ebrahim, S. (2003). Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* 32 (1), 1–22. doi:10.1093/ije/dyg070

Wain, L. V., Shrine, N., Artigas, M. S., Erzurumluoglu, A. M., Noyvert, B., Bossini-Castillo, L., et al. (2017). Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat. Genet.* 49 (3), 416–425. doi:10.1038/ng.3787

Wang, C., Xu, J., Yang, L., Xu, Y., Zhang, X., Bai, C., et al. (2018). Prevalence and risk factors of chronic obstructive pulmonary disease in China (the China pulmonary health [CPH] study): A national cross-sectional study. *Lancet* 391 (10131), 1706–1717. doi:10.1016/S0140-6736(18)30841-9

Wood, A. M., Bassford, C., Webster, D., Newby, P., Rajesh, P., Stockley, R. A., et al. (2011). Vitamin D-binding protein contributes to COPD by activation of alveolar macrophages. *Thorax* 66 (3), 205–210. doi:10.1136/thx.2010.140921

Xie, X., Zhang, Y., Ke, R., Wang, G., Wang, S., Hussain, T., et al. (2015). Vitamin D-binding protein gene polymorphisms and chronic obstructive pulmonary disease susceptibility: A meta-analysis. *Biomed. Rep.* 3 (2), 183–188. doi:10.3892/br.2014.392

Zhou, J. J., Cho, M. H., Castaldi, P. J., Hersh, C. P., Silverman, E. K., and Laird, N. M. (2013). Heritability of chronic obstructive pulmonary disease and related phenotypes in smokers. *Am. J. Respir. Crit. Care Med.* 188 (8), 941–947. doi:10.1164/rccm.201302-0263OC

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 08 November 2022
DOI 10.3389/fgene.2022.1009589

Check for updates

OPEN ACCESS

EDITED BY
Ahmed Rebai,
Centre of Biotechnology of Sfax, Tunisia

REVIEWED BY
Antonio Victor Campos Coelho,
Albert Einstein Israelite Hospital, Brazil
Yong Cui,
China-Japan Friendship Hospital, China
Xianbo Zuo,
China-Japan Friendship Hospital, China

*CORRESPONDENCE
Wenjun Wang,
wangwj0720@sina.com
Liangdan Sun,
ahmusld@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

RECEIVED 02 August 2022
ACCEPTED 18 October 2022
PUBLISHED 08 November 2022

CITATION
Zhang C, Qin Q, Li Y, Zheng X, Chen W, Zhen Q, Li B, Wang W and Sun L (2022), Multifactor dimensionality reduction reveals the effect of interaction between ERAP1 and IFIH1 polymorphisms in psoriasis susceptibility genes. *Front. Genet.* 13:1009589. doi: 10.3389/fgene.2022.1009589

COPYRIGHT
© 2022 Zhang, Qin, Li, Zheng, Chen, Zhen, Li, Wang and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Multifactor dimensionality reduction reveals the effect of interaction between ERAP1 and IFIH1 polymorphisms in psoriasis susceptibility genes

Chang Zhang[1,2,3,4,5†], Qin Qin[1,2,3,4,5†], Yuanyuan Li[1,2,3,4,5†], Xiaodong Zheng[1,2,3,4,5], Weiwei Chen[1,2,3,4,5], Qi Zhen[1,2,3,4,5], Bao Li[1,2,3,4,5], Wenjun Wang[1,2,3,4,5*] and Liangdan Sun[1,2,3,4,5*]

[1]Department of Dermatology, The First Affiliated Hospital of Anhui Medical University, Hefei, China, [2]Institute of Dermatology, Anhui Medical University, Hefei, China, [3]Key Laboratory of Dermatology, Anhui Medical University, Ministry of Education, Hefei, China, [4]Inflammation and Immune Mediated Diseases Laboratory of Anhui Province, Hefei, China, [5]Anhui Provincial Institute of Translational Medicine, Hefei, China

Background: Psoriasis is a common immune-mediated hyperproliferative skin dysfunction with known genetic predisposition. Gene–gene interaction (e.g., between HLA-C and ERAP1) in the psoriasis context has been reported in various populations. As ERAP1 has been recognized as a psoriasis susceptibility gene and plays a critical role in antigen presentation, we performed this study to identify interactions between ERAP1 and other psoriasis susceptibility gene variants.

Methods: We validated psoriasis susceptibility gene variants in an independent cohort of 5,414 patients with psoriasis and 5,556 controls. Multifactor dimensionality reduction (MDR) analysis was performed to identify the interaction between variants significantly associated with psoriasis in the validation cohort and ERAP1 variants. We then conducted a meta-analysis of those variants with datasets from exome sequencing, target sequencing, and validation analyses and used MDR to identify the best gene–gene interaction model, including variants that were significant in the meta-analysis and ERAP1 variants.

Results: We found that 19 of the replicated variants were identified with $p < 0.05$ and detected six single-nucleotide polymorphisms of psoriasis susceptibility genes in the meta-analysis. MDR analysis revealed that the best predictive model was that between the rs27044 polymorphism of ERAP1 and the rs7590692 polymorphism of IFIH1 (cross-validation consistency = 9/10, test accuracy = 0.53, odds ratio = 1.32 (95% CI, 1.09–1.59), $p < 0.01$).

Abbreviations: GWAS, genome-wide association studies; KNN, K-nearest neighbors; MAF, minor allele frequency; MDR, multifactor dimensionality reduction; PCR, polymerase chain reaction; SNP, single-nucleotide polymorphism; SNVs, single-nucleotide variations.

Conclusion: Our findings suggest that the interaction between ERAP1 and IFIH1 affects the development of psoriasis. This hypothesis needs to be tested in basic biological studies.

KEYWORDS

psoriasis, gene–gene interaction, multifactor dimensionality reduction, genome-wide association studies, IFIH1, ERAP1

Introduction

Psoriasis is a common immune-mediated hyperproliferative skin dysfunction with known genetic predisposition that affects almost 125 million people worldwide (Armstrong and Read, 2020). Psoriasis morbidity rates range from 0.09% to 5.1% (Michalek et al., 2017), with major regional differences. The condition results from a combination of interacting genetic, environmental, and immunological factors (Myers et al., 2005; Mahil et al., 2015; Greb et al., 2016). Although progress has been made in the interpretation of its heredity, the pathogenesis of psoriasis remains unclear. Early correlation studies established that the genetic contribution to psoriasis is complex and multifactorial.

Since 2005, genome-wide association studies (GWASs) have been established as the most effective approach to the identification of genetic variations related to complex diseases *via* the use of a complex single-nucleotide polymorphism (SNP) map of the human genome to identify differences in allele frequency between patients and controls (Kruglyak, 2008; Zhang, 2012). Although GWASs have led to the robust identification of numerous susceptibility loci, they capture small proportions of estimated disease heritability; this issue is referred to as the "missing heritability" problem (Manolio et al., 2009; Eichler et al., 2010). Most correlational research involves the detection of SNPs in isolation, without examination of the mutual effects of their interaction, leading intricate reciprocities in the biosystem to be overlooked. With the accumulation of large amounts of genome data, further gene–gene interaction and association studies are essential to explore the biological pathways and pathogeneses of common diseases (Cordell, 2009).

Most common diseases are induced by the nonlinear interaction of numerous genetic and environmental factors. Complex diseases are always generated by gene–environment and gene–gene interactions, the latter of which can enhance their danger (Evans et al., 2006; Cordell, 2009). Many methods can be used to test for gene–gene interaction, including those involving logistic regression models, neural networks, and multifactor dimensionality reduction (MDR) (Cordell, 2009). The use of a traditional (e.g., logistic regression) model to detect multiple interactions at the same time imposes a huge computational burden due to the exponential increase in the number of interaction terms (Chattopadhyay and Lu, 2019). Although neural networks are considered good nonlinear models, they are prone to overfitting (Ritchie et al., 2003). MDR, developed by Ritchie et al. (Ritchie et al., 2001) in 2001, is a non-parametric model-free method that can be applied directly in case-control and discordant sib-pair studies. It involves no *a priori* genetic model assumption, unlike other approaches (e.g., linear regression or generalized linear models); this enables the detection of any genetic interaction in relation to disease. With the division of samples into high- and low-risk groups, MDR enables the reduction of *n*-dimensional models to one-dimension and reduces type-I and-II errors. It is currently the most popular method for detection of gene–gene interaction, and it has been used to identify potential interaction loci in many diseases, including sporadic breast cancer (Ritchie et al., 2001), hypertension (Moore and Williams, 2002), type-2 diabetes mellitus (Cho et al., 2004), and the autoimmune diseases rheumatoid arthritis (Julià et al., 2007; Liu et al., 2011), systemic lupus erythematosus (Zhang et al., 2016), and ankylosing spondylitis (Evans et al., 2011).

Based on existing GWAS data (Zhang et al., 2009), our group has analyzed the interaction of various susceptibility genes (including HLA–IL12B, HLA–ICE (Zheng et al., 2011), HLA-C–ERAP1, and HLA-C–TRAF3IP2 (Yin et al., 2013) interactions) in the Han Chinese population. Such gene interactions, including those related to psoriasis, have been reported in different populations (Strange et al., 2010). ERAP1 has been reported as being related to psoriasis in multiple populations, and our group has genotyped and replicated its polymorphisms (Chen et al., 2022). In this study, we detected interactions of ERAP1 with other psoriasis susceptibility genes in the Han Chinese population.

Materials and methods

Sample collection

This study was conducted with data generated in our previous research—an exome sequencing study conducted with 781 patients with psoriasis and 676 controls and a targeted sequencing study conducted with 9,946 psoriasis cases and 9,906 controls (Tang et al., 2014). We performed a case-control study with an independent sample cohort consisting of 5,414 psoriasis cases and 5,556 controls to validate the psoriasis susceptibility genes. All patients and controls were of Han Chinese origin and had attended the First Affiliated Hospital of Anhui Medical University. The controls were healthy

volunteers without psoriasis, other autoimmune diseases, systemic disease, or a family history of psoriasis (in first-, second-, and third-degree relatives). Each patient was examined by medical professionals and diagnosed with psoriasis by two experienced dermatologists, with rigorous recording of clinical data to ensure their specificity and reliability. After obtaining informed consent, medical professionals collected peripheral blood samples from all patients and controls. This research was approved by the Ethics Committee of Anhui Medical University Committee and performed following Declaration of Helsinki guidelines.

Procedure

Genomic DNA was extracted from the participants' peripheral blood mononuclear cells. DNA concentrations and optical density ratios (A_{260}/A_{280}) were determined using a NanoDrop 1,000 spectrophotometer (Thermo Fisher Scientific, United States). The Sequenom MassArray system (Sequenom, United States) was used for SNP genotyping. We used 15 ng standardized genomic DNA per sample for subsequent genetic typing. We used multiplex polymerase chain reactions to amplify the genomic DNA and then performed site-specific monacyl elongation reactions to obtain products. The amplified products were purified with AgencourtAM SPRI XP microbeads (Beckman Coulter Life Sciences, United States). After desalination, the final products were transferred to a 384-element SpectrCHIP array (Applied Biosystems, United States). Alleles were detected with a matrix-assisted laser desorption ionization time-of-flight 70 mass spectrometer and analyzed using Massarray Typer 71 software (Sequenom, United States). In the independent cohort, validated SNPs were selected using the following terms: position within 500 kb of a psoriasis susceptibility gene tag, $p < 1.0 \times 10^{-4}$ in the exome sequencing and targeted sequencing studies, genotyping detection rate >90%, and Hardy–Weinberg test $p > 1.0 \times 10^{-4}$ in controls.

Gene–gene interaction analysis

Based on our group's previous study (Chen et al., 2022), ERAP1 variants were genotyped and replicated in the present cohort. The statistical analysis was performed with PLINK 1.07 and a significance level of $p < 0.05$. Gene–gene interactions were analyzed with open-source MDR software (ver. 2.0 Beta 2; available from http://sourceforge.net/projects/mdr). As the MDR software requires complete datasets with no missing value, we replaced missing values with numbers that did not represent genotypes; the second method used was *k*-nearest neighbor estimation to fill in the missing values (*k* was equal to the square root of the number of SNPs) (Schwender, 2012).

Meta-analysis of GWAS discovery results

To expand the sample size and detect more gene–gene interactions, we conducted a meta-analysis of summary GWAS statistics from the exome sequencing (n = 1,457), targeted sequencing (n = 19,007), and validation (n = 10,970) cohorts. We performed the analysis with METAL (available from http://www.sph.umich.edu/csg/abecasis/Metal/), with the test statistics weighted by sample size. The genome-wide significance threshold was $p < 5 \times 10^{-8}$. We performed genomic correction of all results to control for potential inflation of the test statistics. Effector allele frequencies were tracked selectively across all files.

Results

Independent cohort validation

Of 39 SNPs selected for genotyping verification, 19 were significant ($p < 0.05$) in the independent cohort validation analysis. IFIH1 showed the strongest correlation with psoriasis (chr_163737871, $p = 6.52 \times 10^{-7}$, odds ratio (OR) = 0.81; chr2_163136771, $p = 1.72 \times 10^{-6}$, OR = 0.81; Table 1). Other validated psoriasis susceptibility genes were ERAP2 (5q15), IL18R1 (2q12.1), LTB (12p13.31), IL1RL1 (2q12.1), CARD14 (17q25.3), and SLC9A4 (2q12.1).

Gene–gene interaction study

The MDR analysis was conducted with the 19 SNPs showing significance in the replicated cohort and six ERAP1 SNPs. The rs27044 polymorphism of ERAP1 was the best single-locus model (CVC = 10, test accuracy = 0.52, OR = 1.32 (95% CI, 1.07–1.63), $p < 0.01$). The best predictive model (i.e., that with maximum testing accuracy) was that between the rs27044 polymorphism of ERAP1 and the rs7590692 polymorphism of IFIH1 (CVC = 9/10, test accuracy = 0.53, OR = 1.32 (95% CI, 1.09–1.59), $p < 0.01$; Table 2). In this model, high-risk genotypes were TT × CC, TC × GG, TC × GC, TC × CC, CC × GC, and CC × CC, and low-risk genotypes were TT × GC, TT × GG, and CC × GG (Figure 1). Figures 2, 3 show measures of characterized epistatic of the models from the interaction analysis.

Results of the meta-analysis and MDR

In the meta-analysis, six of the 39 SNPs reached the genome-wide significance threshold ($p < 10^{-8}$). These SNPs were in IFIH1 (rs12479043, $p = 2.85 \times 10^{-15}$; rs7590692, $p = 1.57 \times 10^{-14}$), CARD14 (rs4889997, $p = 4.03 \times 10^{-12}$), GJB2 (rs72474224, $p = 8.22 \times 10^{-11}$), NFKB1 (rs3817685, $p = 4.24 \times 10^{-10}$), and ERAP2

TABLE 1 SNPs with significance in replication analysis.

Chr.	Variant ID	Gene	F_A	F_U	Allele	*p* value	OR (95% CI)
2q24.3	rs12479043	IFIH1	0.10	0.12	G/C	6.52×10^{-7}	0.81 (0.74–0.88)
2q24.3	rs7590692	IFIH1	0.10	0.12	C/T	1.72×10^{-6}	0.81 (0.74–0.88)
5q15	rs2303208	ERAP2	0.37	0.40	A/G	9.75×10^{-6}	0.88 (0.83–0.93)
2q12.1	rs1882348	IL18R1	0.32	0.35	A/T	2.56×10^{-5}	0.89 (0.84–0.94)
12p13.31	rs12354	LTBR	0.12	0.14	T/G	3.85×10^{-5}	0.85 (0.78–0.92)
2q12.1	rs873022	IL1RL1	0.29	0.32	T/G	4.73×10^{-5}	0.88 (0.83–0.94)
17q25.3	rs4889997	CARD14	0.49	0.51	A/G	5.03×10^{-5}	0.89 (0.85–0.94)
2q12.1	rs61731285	SLC9A4	0.11	0.09	T/C	8.55×10^{-5}	1.20 (1.09–1.31)
2q12.1	rs3213733	IL18R1	0.11	0.10	A/C	9.29×10^{-5}	1.19 (1.09–1.30)
2q12.1	rs12905	IL1RL1	0.30	0.32	A/G	1.03×10^{-4}	0.89 (0.84–0.94)
2q12.1	rs3213732	IL18R1	0.17	0.15	G/A	8.66×10^{-4}	1.14 (1.05–1.23)
2q12.1	rs2287033	IL18R1	0.17	0.15	C/T	1.23×10^{-3}	1.13 (1.05–1.22)
2q12.1	rs6749014	IL18R1	0.16	0.15	T/C	2.18×10^{-3}	1.12 (1.04–1.21)
2q12.1	rs4988956	IL1RL1	0.15	0.13	A/G	5.08×10^{-3}	1.12 (1.03–1.21)
4q24	rs3817685	NFKB1	0.40	0.38	G/C	5.34×10^{-3}	1.08 (1.02–1.14)
1q32.1	rs28694304	C1orf186	0.13	0.12	A/C	8.51×10^{-3}	1.12 (1.03–1.21)
13q12.11	rs72474224	GJB2	0.05	0.05	T/C	2.19×10^{-2}	1.15 (1.02–1.30)
19p13.2	rs280497	TYK2	0.42	0.41	A/G	2.78×10^{-2}	1.06 (1.01–1.12)
17q25.3	rs2066964	CARD14	0.46	0.47	C/G	3.68×10^{-2}	0.94 (0.89–1.00)

Chr. chromosome, F_A frequency in cases, F_U frequency in controls.

TABLE 2 Best multifactor dimensionality reduction (MDR) interaction models.

Locus number	Number of the risk factors (best interaction model)	Testing accuracy	CVC	0R	95% CI	*p*
1	ERAP1_rs27044	0.5231	10/10	1.32	1.06–1.64	0.0098
2	IFIH1_rs7590692,ERAP1_rs27044	0.5314	9/10	1.32	1.09–1.59	0.0036
3	IFIH1_rs7590692,ERAP1_rs27044,CARD14_rs4889997	0.5265	6/10	1.25	1.04–1.49	0.0147

The model with the maximum testing accuracy was considered the best model. MDR, multifactor dimensionality reduction; CVC, cross-validation consistency.

(rs2303208, $p = 7.50 \times 10^{-10}$; Table 3). In the MDR analysis of these SNPs and six SNPs of ERAP1, the best model was still that between the rs27044 polymorphism of ERAP1 and the rs7590692 polymorphism of IFIH1 (CVC = 9/10, test accuracy = 0.53, OR = 1.32 (95% CI, 1.09–1.59) $p < 0.01$).

Discussion

With the deep development of psoriasis-related genetic analysis, our group has generated systemic genomic data on psoriasis through exome and targeted sequencing. In this study, we verified the psoriasis susceptibility genes IFIH1 (2q24.3), ERAP2 (5q15), IL18R1 (2q12.1), LTBR (12p13.31), IL1RL1 (2q12.1), CARD14 (17q25.3), and SLC9A4 (2q12.1). We found that the most significant interaction was between the rs27044 polymorphism of ERAP1 and the rs7590692 of IFIH1; this result was confirmed in the MDR analysis.

ERAP1 is an immune-related gene encoding multifunctional aminopeptidase that is induced by interferon-γ. It plays critical roles in antigen presentation and N-terminal peptide pruning to reach the best presenting size of the major histocompatibility complex I molecule. These molecules can present polypeptides containing 8–10 amino acids directly outside of cells, enhancing antigen presentation and T-cell activation in disease processes (Kochan et al., 2011). ERAP1 is more likely to be present in individuals carrying the HLA-C susceptibility allele, with the first and most important psoriasis susceptibility gene identified (Bowcock, 2005). A recent study revealed that ERAP1 causes psoriasis by affecting HLA-C production *via* melanocyte autoantigens (Arakawa et al., 2021). Thus, it may cause psoriasis lesions by interacting with HLA-C.

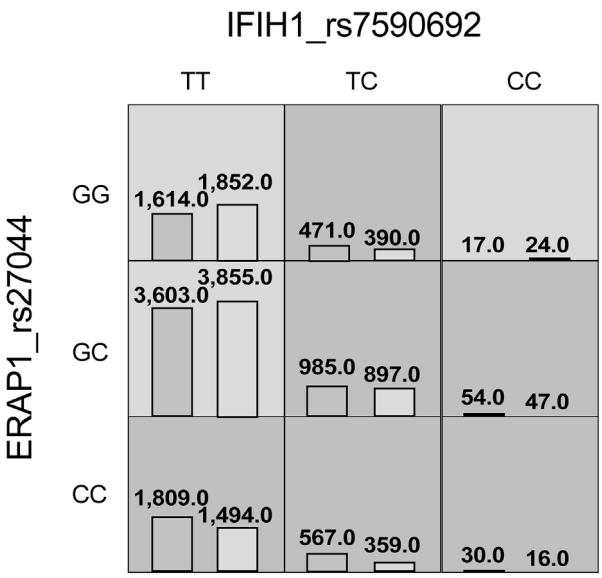


FIGURE 1
High- and low-risk genotypes in the best psoriasis interaction model. Dark-gray cells represent high risk (combinations exceeding the *r* ratio of the percentage of cases to controls ≥1.0) and light-gray cells represent low risk (combinations below the ratio threshold).

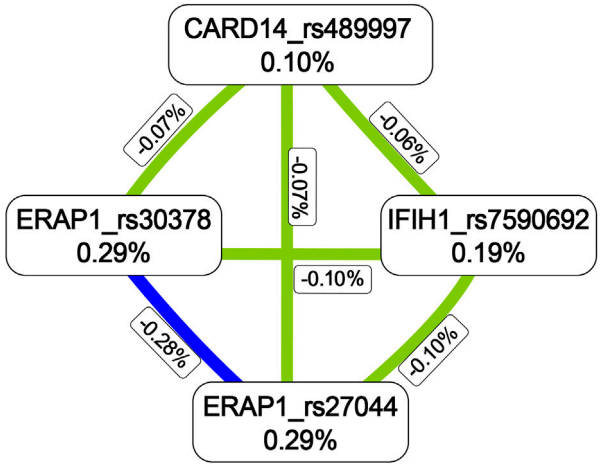


FIGURE 2
Circular graph of the best models. The numbers between two SNPs was information gain, which was used to detect and characterize epistatics of the interaction model and calculated by constructing two attributes using the MDR function. Red: positive information gain (synergistic or non-additive effect); blue: negative information gain (redundancy or correlation); yellow: no information gain (independence).

ERAP1 has been related to psoriasis in Chinese (rs151823) (Sun et al., 2010) and European (rs27524) (Strange et al., 2010) populations. In a meta-analysis of nine case-control studies, the rs27044 polymorphism of ERAP1 was associated significantly with psoriasis (Wu and Zhao, 2021). This gene has also been associated with ankylosing spondylitis (Tsui et al., 2010), Behcet's disease, multiple sclerosis, and inflammatory bowel disease (Reeves and James, 2018).

We replicated IFIH1 in our independent sample cohort with $p = 6.52 \times 10^{-7}$. IFIH1 encodes MDA5, an innate pattern recognition receptor. MDA5 is a cytoplasmic sensor of picornavirus nucleic acids that triggers a cascade of antiviral responses, including the induction of type I interferons and pro-inflammatory cytokines. Once it is bound to pathogen-like viral RNA, activated MDA5 interacts with mitochondrial antiviral-signaling proteins and then phosphorylated interferon regulatory factors (IRFs) 3 and 7. Dimers are formed by phosphorylated IRF-3 and -7 and move into cell nuclei to stimulate the secretion of type 1 interferon (IFN1) (Brisse and Ly, 2019). IFN1 mediates autoimmune responses to invaders such as viruses. MDA5 has been observed to significantly increase psoriatic plaques and keratinocytes, and its overexpression acts as a negative regulator for the differentiation of calcium-induced keratinocytes in psoriatic skin lesions (Hong et al., 2021). IFIH1 has been related to psoriasis in a European population (rs17716942) (Strange et al., 2010), and its rs35667974 and rs10930046 polymorphisms were found to have protective effects against psoriasis in a white North American population (Li et al., 2010). In addition, evidence supports the involvement of IFIH1 SNPs in inflammatory diseases such as type 1 diabetes (Nejentsev et al., 2009), inflammatory bowel disease (Cananzi et al., 2021), and dermatomyositis (Kurtzman and Vleugels, 2018).

Our MDR analysis revealed a significant association between ERAP1 and IFIH1, suggesting that the presence of the ERAP1 rs27044 polymorphism affects the expression of the IFIH1 rs7590692 polymorphism. In an *in vitro* experiment (Aldhamen et al., 2013), increased levels of proinflammatory cytokines such as tumor necrosis factor (TNF)-α were observed in ERAP1-knockout mice, suggesting that the lack of ERAP1 expression induces increased autoimmune activity. TNF-α, a cytokine commonly elevated in patients with psoriasis, promotes the transcription of MDA5 and RIG-I in keratinocytes (Kitamura et al., 2007; Racz et al., 2011). Taken together, these molecular biological findings suggest that ERAP1 might regulate the expression of IFIH1, but more basic biological studies are needed to determine how this gene interaction affects psoriasis.

This study has several limitations. The results provide evidence of correlations between gene–gene interactions and psoriasis but no information about the mechanisms of resulting effects, and the accuracy of the best model was not high. We will carry out further basic biological studies to verify this result. In addition, the study was conducted with data from a Han Chinese population; similar studies need to be conducted with data from other ethnic groups.

In summary, we screened for interactions between genes that affect psoriasis susceptibility and found that the most significant interaction was between the rs27044 polymorphism of ERAP1 and the rs7590692 polymorphism of IFIH1. Psoriasis

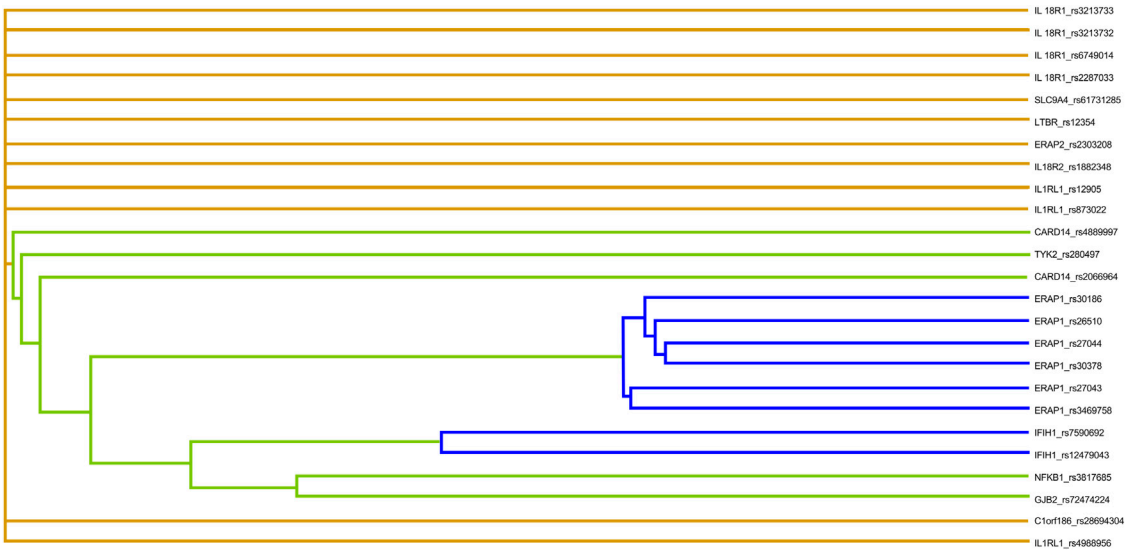


FIGURE 3
Dendrogram between polymorphisms in psoriasis susceptibility genes. Colors depict degrees of synergy, ranging from red (most information gain) to yellow–green and blue (most information redundancy).

TABLE 3 Result of meta-analysis.

Variant ID	Gene	Chr.	A1	A2	Freq1	P_{Meta}	Direction	I^2	P_{Het}
rs12479043	IFIH1	2q24.3	C	G	0.88	2.85×10^{-15}	+++	0	0.91
rs7590692	IFIH1	2q24.3	T	C	0.88	1.57×10^{-14}	-++	7.1	0.34
rs4889997	CARD14	17q25.3	A	G	0.50	4.03×10^{-12}	---	65.8	0.05
rs72474224	GJB2	13q12.11	T	C	0.05	8.22×10^{-11}	+++	59.9	0.08
rs3817685	NFKB1	4q24	C	G	0.63	4.24×10^{-10}	?--	34.5	0.22
rs2303208	ERAP2	5q15	A	G	0.40	7.50×10^{-10}	+--	21.5	0.28

Chr.: chromosome; A1: minor allele; A2: major allele; Freq1: frequency of allele 1; P_{Meta}: *p* value of the meta-analysis; P_{Het}: *p* value of heterogeneity.

is a multifactorial autoimmune disease with many genetic contributions. Although our comprehension of the genetic mechanisms underlying it has increased and the interaction of susceptibility genes has been identified based on SNPs, many questions remain. A better understanding of how these gene interactions affect the development of psoriasis, generated from basic biological studies, may provide answers to these questions.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of there pository/repositories and accession number(s) can be found below: Zenodo with accession 7260050—(https://doi.org/10.5281/zenodo.7260050).

Ethics statement

The studies involving human participants were reviewed and approved by the Clinical Research Ethics Committee of Anhui Medical University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

LS and WW conceived and designed the research. CZ wrote this manuscript. QQ and YL were responsible for selecting blood samples and extracting DNA. CZ, QQ, and YL contributed equally to this work. XZ, WC, BL, and QZ performed data

analysis. All authors contributed to the study and approved the final manuscript.

Funding

This study was supported by grants from the Anhui Institute of Translation Medicine (ZHY 2020A005) and the University Synergy Innovation Program of Anhui Province (GXXT-2020-064).

Acknowledgments

We thank the patients and their families who contributed to this study.

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.

References

Aldhamen, Y. A., Seregin, S. S., Rastall, D. P. W., Aylsworth, C. F., Pepelyayeva, Y., and Busuito, C. J., (2013). Endoplasmic reticulum aminopeptidase-1 functions regulate key aspects of the innate immune response. *PloS one* 8 (7), e69539. doi:10.1371/journal.pone.0069539

Arakawa, A., Reeves, E., Vollmer, S., Arakawa, Y., He, M., Galinski, A., et al. (2021). ERAP1 controls the autoimmune response against melanocytes in psoriasis by generating the melanocyte autoantigen and regulating its amount for HLA-C*06:02 presentation. *J. Immunol.* 207 (9), 2235–2244. doi:10.4049/jimmunol.2100686

Armstrong, A. W., and Read, C. (2020). Pathophysiology, clinical presentation, and treatment of psoriasis: A review. *JAMA* 323 (19), 1945–1960. doi:10.1001/jama.2020.4006

Bowcock, A. M. (2005). The genetics of psoriasis and autoimmunity. *Annu. Rev. Genomics Hum. Genet.* 6, 93–122. doi:10.1146/annurev.genom.6.080604.162324

Brisse, M., and Ly, H. (2019). Comparative structure and function analysis of the RIG-I-like receptors: RIG-I and MDA5. *Front. Immunol.* 10, 1586. doi:10.3389/fimmu.2019.01586

Cananzi, M., Wohler, E., Marzollo, A., Colavito, D., You, J., and Jing, H., (2021). IFIH1 loss-of-function variants contribute to very early-onset inflammatory bowel disease. *Hum. Genet.* 140 (9), 1299–1312. doi:10.1007/s00439-021-02300-4

Chattopadhyay, A., and Lu, T. P. (2019). Gene-gene interaction: The curse of dimensionality. *Ann. Transl. Med.* 7 (24), 813. doi:10.21037/atm.2019.12.87

Chen, W., Yong, L., Ge, H., Xu, Q., Zhen, Q., and Li, B., (2022). Polymorphisms in ERAP1 gene are associated with psoriasis. *Meta Gene* 31, 100995. doi:10.1016/j.mgene.2021.100995

Cho, Y. M., Ritchie, M. D., Moore, J. H., Park, J. Y., Lee, K. U., and Shin, H. D., (2004). Multifactor-dimensionality reduction shows a two-locus interaction associated with Type 2 diabetes mellitus. *Diabetologia* 47 (3), 549–554. doi:10.1007/s00125-003-1321-3

Cordell, H. J. (2009). Detecting gene-gene interactions that underlie human diseases. *Nat. Rev. Genet.* 10 (6), 392–404. doi:10.1038/nrg2579

Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., and Moore, J. H., (2010). Missing heritability and strategies for finding the underlying causes of complex disease. *Nat. Rev. Genet.* 11 (6), 446–450. doi:10.1038/nrg2809

Evans, D. M., Marchini, J., Morris, A. P., and Cardon, L. R. (2006). Two-stage two-locus models in genome-wide association. *PLoS Genet.* 2 (9), e157. doi:10.1371/journal.pgen.0020157

Evans, D. M., Spencer, C. C. A., Pointon, J. J., Su, Z., Harvey, D., and Kochan, G., (2011). Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nat. Genet.* 43 (8), 761–767. doi:10.1038/ng.873

Greb, J. E., Goldminz, A. M., Elder, J. T., Lebwohl, M. G., Gladman, D. D., and Wu, J. J., (2016). *Nat. Rev. Dis. Prim.* 2, 16082. doi:10.1038/nrdp.2016.82

Hong, D. K., Choi, M. R., Hwang, Y. L., Lee, J. K., Lee, Y., and Seo, Y. J., (2021). Potential role of cytosolic RNA sensor MDA5 as an inhibitor for keratinocyte differentiation in the pathogenesis of psoriasis. *Ann. Dermatol.* 33 (4), 339–344. doi:10.5021/ad.2021.33.4.339

Julià, A., Moore, J., Miquel, L., Alegre, C., Barcelo, P., and Ritchie, M., (2007). Identification of a two-loci epistatic interaction associated with susceptibility to rheumatoid arthritis through reverse engineering and multifactor dimensionality reduction. *Genomics* 90 (1), 6–13. doi:10.1016/j.ygeno.2007.03.011

Kitamura, H., Matsuzaki, Y., Kimura, K., Nakano, H., Imaizumi, T., and Satoh, K., (2007). Cytokine modulation of retinoic acid-inducible gene-I (RIG-I) expression in human epidermal keratinocytes. *J. Dermatol. Sci.* 45 (2), 127–134. doi:10.1016/j.jdermsci.2006.11.003

Kochan, G., Krojer, T., Harvey, D., Fischer, R., Chen, L., and Vollmar, M., (2011). Crystal structures of the endoplasmic reticulum aminopeptidase-1 (ERAP1) reveal the molecular basis for N-terminal peptide trimming. *Proc. Natl. Acad. Sci. U. S. A.* 108 (19), 7745–7750. doi:10.1073/pnas.1101262108

Kruglyak, L. (2008). The road to genome-wide association studies. *Nat. Rev. Genet.* 9 (4), 314–318. doi:10.1038/nrg2316

Kurtzman, D. J. B., and Vleugels, R. A. (2018). Anti-melanoma differentiation-associated gene 5 (MDA5) dermatomyositis: A concise review with an emphasis on distinctive clinical features. *J. Am. Acad. Dermatol.* 78 (4), 776–785. doi:10.1016/j.jaad.2017.12.010

Li, Y., Liao, W., Cargill, M., Chang, M., Matsunami, N., and Feng, B. J., (2010). Carriers of rare missense variants in IFIH1 are protected from psoriasis. *J. Invest. Dermatol.* 130 (12), 2768–2772. doi:10.1038/jid.2010.214

Liu, C., Ackerman, H. H., and Carulli, J. P. (2011). A genome-wide screen of gene-gene interactions for rheumatoid arthritis susceptibility. *Hum. Genet.* 129 (5), 473–485. doi:10.1007/s00439-010-0943-z

Mahil, S. K., Capon, F., and Barker, J. N. (2015). Genetics of psoriasis. *Dermatol. Clin.* 33 (1), 1–11. doi:10.1016/j.det.2014.09.001

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., and Hunter, D. J., (2009). Finding the missing heritability of complex diseases. *Nature* 461 (7265), 747–753. doi:10.1038/nature08494

Michalek, I. M., Loring, B., and John, S. M. (2017). A systematic review of worldwide epidemiology of psoriasis. *J. Eur. Acad. Dermatol. Venereol.* 31 (2), 205–212. doi:10.1111/jdv.13854

Moore, J. H., and Williams, S. M. (2002). New strategies for identifying gene-gene interactions in hypertension. *Ann. Med.* 34 (2), 88–95. doi:10.1080/07853890252953473

Myers, A., Kay, L. J., Lynch, S. A., and Walker, D. J. (2005). Recurrence risk for psoriasis and psoriatic arthritis within sibships. *Rheumatol. Oxf.* 44(6): p. 773–776. doi:10.1093/rheumatology/keh589

Nejentsev, S., Walker, N., Riches, D., Egholm, M., and Todd, J. A. (2009). Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* 324 (5925), 387–389. doi:10.1126/science.1167728

Racz, E., Prens, E. P., KantM., FlorEncia, E., Jaspers, N. G., and Laman, J. D., (2011). Narrowband ultraviolet B inhibits innate cytosolic double-stranded RNA receptors in psoriatic skin and keratinocytes. *Br. J. Dermatol.* 164 (4), 838–847. doi:10.1111/j.1365-2133.2010.10169.x

Reeves, E., and James, E. (2018). The role of polymorphic ERAP1 in autoinflammatory disease. *Biosci. Rep.* 38 (4), BSR20171503. doi:10.1042/BSR20171503

Ritchie, M. D., Hahn, L. W., RoodiN., Bailey, L. R., Dupont, W. D., Parl, F. F., et al. (2001). Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am. J. Hum. Genet.* 69 (1), 138–147. doi:10.1086/321276

Ritchie, M. D., White, B. C., Parker, J. S., Hahn, L. W., and Moore, J. H. (2003). Optimization of neural network architecture using genetic programming improves detection and modeling of gene-gene interactions in studies of human diseases. *BMC Bioinforma.* 4, 28. doi:10.1186/1471-2105-4-28

Schwender, H. (2012). Imputing missing genotypes with weighted k nearest neighbors. *J. Toxicol. Environ. Health. A* 75 (8-10), 438–446. doi:10.1080/15287394.2012.674910

Strange, A., Capon, F., Spencer, C. C. A., Knight, J., Weale, M. E., Allen, M. H., et al. (2010). A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. *Nat. Genet.* 42 (11), 985–990. doi:10.1038/ng.694

Sun, L.-D., Cheng, H., Wang, Z. X., Zhang, A. P., Wang, P. G., and Xu, J. H., (2010). Association analyses identify six new psoriasis susceptibility loci in the Chinese population. *Nat. Genet.* 42 (11), 1005–1009. doi:10.1038/ng.690

Tang, H., Jin, X., Li, Y., Jiang, H., Tang, X., and Yang, X., (2014). A large-scale screen for coding variants predisposing to psoriasis. *Nat. Genet.* 46 (1), 45–50. doi:10.1038/ng.2827

Tsui, F. W. L., Haroon, N., Reveille, J. D., Rahman, P., Chiu, B., and Tsui, H. W., (2010). Association of an ERAP1 ERAP2 haplotype with familial ankylosing spondylitis. *Ann. Rheum. Dis.* 69 (4), 733–736. doi:10.1136/ard.2008.103804

Wu, X., and Zhao, Z. (2021). Associations between ERAP1 gene polymorphisms and psoriasis susceptibility: A meta-analysis of case-control studies. *Biomed. Res. Int.* 2021, 5515868. doi:10.1155/2021/5515868

Yin, X. Y., Zhang, R., Cheng, H., Pan, Q., Shen, C. B., and Fan, X., (2013). Gene-gene interactions between HLA-C, ERAP1, TNFAIP3 and TRAF3IP2 and the risk of psoriasis in the Chinese Han population. *Br. J. Dermatol.* 169 (4), 941–943. doi:10.1111/bjd.12442

Zhang, X. (2012). Genome-wide association study of skin complex diseases. *J. Dermatol. Sci.* 66 (2), 89–97. doi:10.1016/j.jdermsci.2012.02.017

Zhang, X. J., Huang, W., Yang, S., Sun, L. D., Zhang, F. Y., and Zhu, Q. X., (2009). Psoriasis genome-wide association study identifies susceptibility variants within LCE gene cluster at 1q21. *Nat. Genet.* 41 (2), 205–210. doi:10.1038/ng.310

Zhang, Y., Yang, J., Zhang, J., Sun, L., Hirankarn, N., and Pan, H. F., (2016). Genome-wide search followed by replication reveals genetic interaction of CD80 and ALOX5AP associated with systemic lupus erythematosus in Asian populations. *Ann. Rheum. Dis.* 75 (5), 891–898. doi:10.1136/annrheumdis-2014-206367

Zheng, H. F., Zuo, X. B., Lu, W. S., Li, Y., Cheng, H., and Zhu, K. J., (2011). Variants in MHC, LCE and IL12B have epistatic effects on psoriasis risk in Chinese population. *J. Dermatol. Sci.* 61 (2), 124–128. doi:10.1016/j.jdermsci.2010.12.001

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 02 March 2023
DOI 10.3389/fgene.2023.1114601

Check for updates

OPEN ACCESS

EDITED BY
Xianyong Yin,
Department of Biostatistics, University of Michigan, United States

REVIEWED BY
John Ding,
University of Alabama, United States
Farida Ahangari,
Yale University, United States

*CORRESPONDENCE
Yiju Cheng,
chengchengyiju@126.com
Wenting Yang,
807580186@qq.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

RECEIVED 02 December 2022
ACCEPTED 20 February 2023
PUBLISHED 02 March 2023

CITATION
Fu C, Chen L, Cheng Y, Yang W, Zhu H, Wu X and Cai B (2023), Identification of immune biomarkers associated with basement membranes in idiopathic pulmonary fibrosis and their pan-cancer analysis.
Front. Genet. 14:1114601.
doi: 10.3389/fgene.2023.1114601

COPYRIGHT
© 2023 Fu, Chen, Cheng, Yang, Zhu, Wu and Cai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Identification of immune biomarkers associated with basement membranes in idiopathic pulmonary fibrosis and their pan-cancer analysis

Chenkun Fu[1†], Lina Chen[2,3†], Yiju Cheng[4,3*], Wenting Yang[1*], Honglan Zhu[1], Xiao Wu[1] and Banruo Cai[5]

[1]Department of Respiratory and Critical Care Medicine, The Affiliated Hospital of Guizhou Medical University, Guiyang, China, [2]Guiyang Public Health Clinical Center, Guiyang, China, [3]Guizhou Medical University, Guiyang, China, [4]Department of Respiratory and Critical Care Medicine, The First People's Hospital of Guiyang, Guiyang, China, [5]Shanghai Institute of Technology, Shanghai, China

Idiopathic pulmonary fibrosis (IPF) is a chronic progressive interstitial lung disease of unknown etiology, characterized by diffuse alveolitis and alveolar structural damage. Due to the short median survival time and poor prognosis of IPF, it is particularly urgent to find new IPF biomarkers. Previous studies have shown that basement membranes (BMs) are associated with the development of IPF and tumor metastasis. However, there is still a lack of research on BMs-related genes in IPF. Therefore, we investigated the expression level of BMs genes in IPF and control groups, and explored their potential as biomarkers for IPF diagnosis. In this study, the GSE32537 and GSE53845 datasets were used as training sets, while the GSE24206, GSE10667 and GSE101286 datasets were used as validation sets. In the training set, seven immune biomarkers related to BMs were selected by differential expression analysis, machine learning algorithm (LASSO, SVM-RFE, Randomforest) and ssGSEA analysis. Further ROC analysis confirmed that seven BMs-related genes played an important role in IPF. Finally, four immune-related Hub genes (*COL14A1*, *COL17A1*, *ITGA10*, *MMP7*) were screened out. Then we created a logistic regression model of immune-related hub genes (IHGs) and used a nomogram to predict IPF risk. The nomogram model was evaluated to have good reliability and validity, and ROC analysis showed that the AUC value of IHGs was 0.941 in the training set and 0.917 in the validation set. Pan-cancer analysis showed that IHGs were associated with prognosis, immune cell infiltration, TME, and drug sensitivity in 33 cancers, suggesting that IHGs may be potential targets for intervention in human diseases including IPF and cancer.

KEYWORDS

idiopathic pulmonary fibrosis, basement membrane, immune, pan-cancer, interstitial lung disease

1 Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic progressive interstitial lung disease of unknown etiology (Richeldi et al., 2017). Its pathological features are diffuse alveolitis and alveolar structural damage, eventually forming honeycomb lung (Wolters et al., 2014). The clinical symptoms of IPF include dry cough, fatigue, and progressive exertional dyspnea. IPF is a rare

disease, but its incidence is increasing and is more common in elderly male patients (Maher et al., 2021; Kondoh et al., 2022). Although disease progression varies greatly among individuals, the median survival after diagnosis is less than 3–5 years. Currently, treatment options for patients with IPF remain limited. Anti-fibrotic drugs (Pirfenidone and Nintedanib) have been approved to treat IPF, but they only slow the decline in lung function in IPF patients and do not improve survival (King et al., 2014; Richeldi et al., 2014). These drugs can also cause gastrointestinal adverse reactions, which limits their widespread use to some extent (Bargagli et al., 2019). At present, lung transplantation remains the only effective treatment. Unfortunately, not all patients are suitable for transplant, and complications following the transplant place a huge burden on patients.

Basement membranes (BMs) are a specialized form of extracellular matrix found in various organs of the human body, providing structural support for epithelium, endothelium, muscle, adipocytes, schwann cells, and axons (Mak and Mei, 2017). BMs are mainly composed of laminin and type IV collagen, which are linked together by nicolysaccharide and heparan sulfate proteoglycans to form different BMs in various tissues (Kruegel and Miosge, 2010). BMs can direct cell polarity, differentiation, migration and survival. For example, BMs can control epithelial growth and differentiation during embryonic development (Kyprianou et al., 2020). In cancer, the breakdown of the basement membrane promotes metastasis (Banerjee et al., 2022). Variations in the BMs gene are closely associated with many human diseases (Horton and Barrett, 2021; Wilson, 2022). BMs can mark the pathway of cell migration and epithelialization during tissue repair (Vracko, 1974; Rousselle et al., 2019). Changes in BMs homeostasis can lead to abnormal ECM aggregation and fibrosis (Wilson, 2020). Alveolar BMs promote gas exchange between alveoli and capillaries and regulate the function of cytokines and growth factors (West and Mathieu-Costello, 1999). The integrity of BMs maintains the normal lung structure and is critical for restoring alveolar epithelial homeostasis after lung injury (Strieter and Mehrad, 2009). However, a loss of alveolar and capillary BMs integrity was observed in IPF, suggesting that BMs are involved in IPF genesis (Chen et al., 2016).

Increasing evidence supports the important role of immune response in IPF. On the one hand, damage to lung epithelial cells leads to the production of pro-inflammatory cytokines such as IL-1 and IL-6 in M1 alveolar macrophages (Huang et al., 2018). These cytokines play an important role in host resistance to pathogen invasion. Inhibition of TNF-α secretion can alleviate bleomycin-induced pulmonary fibrosis and collagen deposition (Matsuhira et al., 2020). On the other hand, under chronic inflammatory conditions, Th2 cells secrete cytokines to gradually transform pro-inflammatory M1 macrophages into pro-fibrotic M2 macrophages (Shapouri-Moghaddam et al., 2018). M2 macrophages secrete multiple chemokines and activate Wnt/β-catenin signaling pathways leading to fibroblast activation, myofibroblast differentiation and extracellular matrix remodeling (Wynn and Vannella, 2016). Inhibition of Wnt/β-catenin signaling attenuates M2 macrophage-induced myofibroblast differentiation and bleomycin-induced pulmonary fibrosis (Hou et al., 2018). Another study found that vaccination inhibited M2 macrophage production and fibrocyte recruitment in bleomycin-induced pulmonary fibrosis (Collins et al., 2012). Moreover, the importance of immune responses in IPF has been confirmed by genetic studies, such as DEP domain containing MTOR interacting protein (*DEPTOR*) increase the risk or susceptibility of IPF(Allen et al., 2020).

In this article, we aim to explore the immune markers associated with BMs in IPF and construct a nomogram model to predict the risk of IPF in patients. Increased evidence suggests that IPF is closely linked to cancer (Ballester et al., 2019; Tzouvelekis et al., 2019). However, little is known about the relationship between IPF and cancer. Therefore, we conducted an in-depth analysis of the role of BMs-related immune biomarkers in pan-cancer to explore the common pathogenesis of IPF and cancer, and to find potential therapeutic targets for patients with IPF and cancer. In Figure 1, you can see the workflow chart.

2 Materials and methods

2.1 Data download and processing

We obtained 222 BMs related genes from previous studies (Jayadev et al., 2022). We obtained datasets numbered GSE32537, GSE53845, GSE24206, GSE10667 and GSE101286 from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). The GSE32537, GSE53845 datasets were used as training set. Datasets GSE24206, GSE10667 and GSE101286 were used as validation sets for independent external validation. The main features of the five datasets were shown in Table 1. Datasets were merged after identity document transformation. The original data was normalized through R software "sva" package, followed by classification of the datasets into two categories: IPF and control groups. Finally, we evaluated the quality of the dataset using principal component analysis and plotted the PCA plot *via* R software "ggplot2" package.

2.2 Identification of differentially expressed genes in BMs

We screened differentially expressed genes (DEGs) through R software "limma" package, filter condition for $p < 0.05$ and $|\log_2 FC| > 1$. A volcano map was drawn through the "ggplot2" package of R software to show DEGs. The "VennDiagram" package in R was used to obtain the intersection of BMs-related genes and DEGs, and the differentially expressed genes of basement membranes (BMDEGs) were obtained.

2.3 Protein–protein interaction (PPI) and enrichment analysis of BMDEGs

We visualized the PPI network of BMDEGs *via* the STRING database and performed GO, KEGG, and DO enrichment analyses of BMDEGs *via* the "ClusterProfiler" and "DOSE" packages of R software.

2.4 Three machine learning algorithms for screening disease candidate genes

We used LASSO, SVM-RFE and RandomForest machine learning algorithms to screen disease candidate genes. LASSO

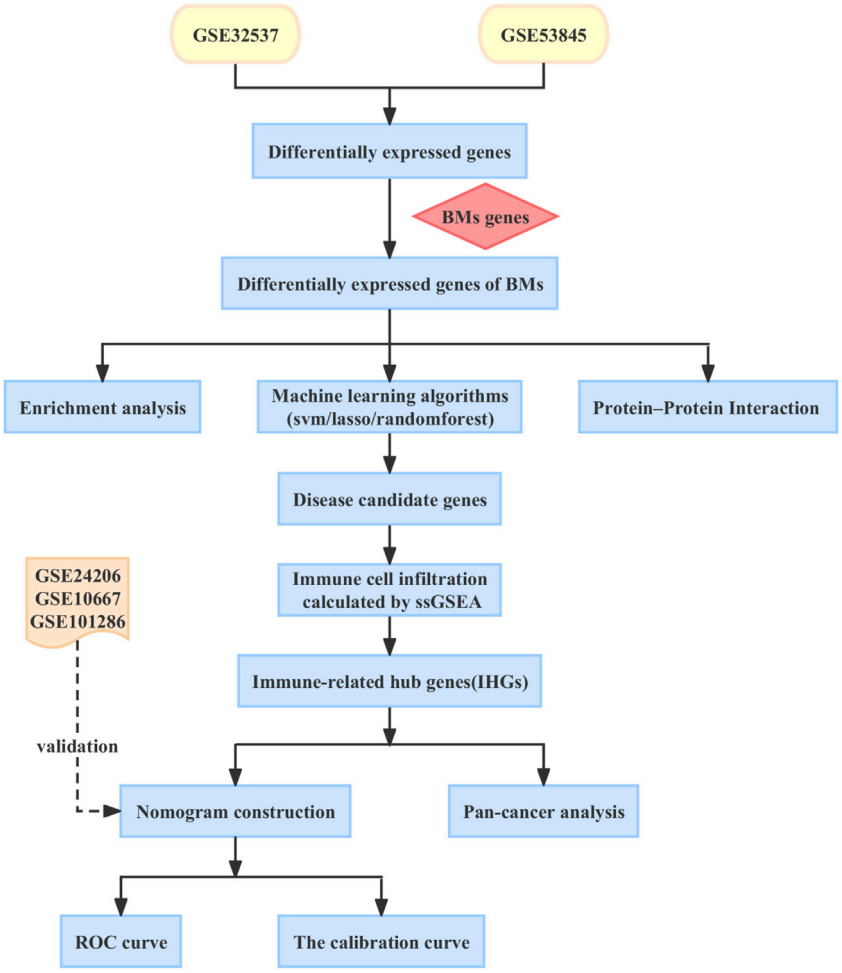


FIGURE 1
The workflow chart of our study.

analysis was performed using 10-fold cross-validated penalty parameters *via* the "glmnet" package of R software. The minimal bimomial deviation was used to determine the optimal penalty parameter lambda. SVM-RFE algorithm detects the points with the minimum cross-validation error through the "e1071", "kernlab" and "caret" packages in R software to screen disease candidate genes. The RandomForest algorithm uses the "randomforest" package of R software to screen disease candidate genes. The Venn diagram visualizes disease candidate genes obtained from the results of three machine learning algorithms.

2.5 Validation of disease candidate genes

To understand the specificity and sensitivity of disease candidate genes for IPF diagnosis, we draw the receiver operating characteristic (ROC) curve through the "pROC" package of R software. Results were presented in the form of area under the curve (AUC). If the AUC of candidate genes was greater than 0.6, we believed that it had diagnostic significance for IPF. Disease candidate gene expression in the IPF and control groups was shown in the box plot.

2.6 Analysis of immune infiltration

We performed correlation analysis of immunity through the "corrplot" package of R software and plotted the correlation heatmap. R software "ggpubr" and "reshape2" packages were used to analyze the differential expression of immune cells and immune functions in IPF and control groups. Spearman correlation analysis was conducted through the "psych" and "ggcorrplot" packages of the R software to analyze the correlation between disease candidate genes and immunity. The screening criteria for IPF immune-related hub genes (IHGs) were more than 1/2 immune infiltration and the correlation coefficient was greater than 0.2.

2.7 Establishment and validation of IHGs risk model

We constructed the nomogram of IHGs and plotted the calibration curve to determine the reliability of the nomogram through the "rms" package of R software. ROC curves were

TABLE 1 An overview of the main features of the dataset used in this study.

Dataset	Platform	IPF	Normal	Publication years	Used for
GSE32537	GPL6244	167	50	2013	DEGs analysis
GSE53845	GPL6480	40	8	2014	DEGs analysis
GSE24206	GPL570	17	6	2011	Model validation
GSE10667	GPL4133	31	15	2009	Model validation
GSE101286	GPL6947	12	3	2017	Model validation

plotted to assess the accuracy of IHGs in diagnosing IPF *via* the "ROCR" package of R software.

2.8 Differential analysis of IHGs in human cancer

We downloaded tumor transcriptome data, clinical data, immune subtype data, mutation data and stemness score (RNAss and DNAss) from the UCSC Xena database (https://xenabrowser.net/). The expression levels of IHGs in 33 tumor and adjacent samples were extracted using the "limma" package of R software. Then, we retained tumor types with more than 5 paracancer tissues, and analyzed the expression of IHGs in tumors and paracancer tissues through the "ggpubr" package of R software.

2.9 Survival analysis of IHGs in pan-cancer

COX regression analysis was conducted through the "survival" package of R software to determine whether IHGs expression was correlated with the survival time and survival status of cancer patients, and the results were presented in forest plots. Survival analysis was performed through the "survival" and "surminer" packages of R software to determine whether IHGs expression was linked with the prognosis of tumor patients.

2.10 Mutation analysis of IHGs in pan-cancer

Tumor mutation burden (TMB) and microsatellite instability (MSI) is specific indicators for predicting immunotherapy in cancer patients. Therefore, we used the "fmsb" package of R software to plot radar maps of TMB and MSI to determine the correlation between IHGs expression and 33 types of tumors.

2.11 Tumor microenvironment (TME) and tumor stemness analysis

We calculated the immune score, stromal score, and estimated score for each sample in the tumor using the "estimate" package of R software. The correlation between IHGs expression and purity in 33 tumors was analyzed by Spearman correlation analysis. The relationship between IHGs expression and tumor stemness was determined by the "corrplot" package of R software.

2.12 Immune analysis of IHGs in pan-cancer

Immune subtypes including C1(Wound Healing), C2(IFN γ Dominant), C3(Inflammatory), C4 (lymphocyte Depleted), C5 (M2 macrophages Dominant), C6 (TGF-β Dominant) subtypes. Previous studies have shown that among the six immune subtypes, C4 and C6 are associated with lower survival rates, while C3 and C5 are the opposite (Tamborero et al., 2018). The relationship between IHGs expression and immune subtypes was analyzed by the "limma", "ggplot2" and "reshape2" packages of R software. The correlation between IHGs expression and immune checkpoints was completed by Pearson correlation analysis. Finally, we investigated the correlation between IHGs expression in tumors and 21 types of immune cells using TIMER2.0 database (http://timer.cistrome.org/).

2.13 Enrichment analysis

We used the GeneMANIA database (http://genemania.org/) to predict and visualize genes that function similar to IHGs. The Metascape website (https://metascape.org/) was used to analyze the functions in which genes may be involved.

2.14 Drug sensitivity analysis of IHGs in pan-cancer

We downloaded drug sensitivity data for 60 human cancers from the CellMiner websites (https://discover.nci.nih.gov/cellminer/) and screened 263 FDA-approved or clinical trial drugs for this study. The relationship between IHGs expression and drugs was analyzed by Pearson correlation analysis.

2.15 Statistical analysis

Statistical tests were performed using R software (version 4.1.3). For all statistical analyses, $p < 0.05$ was considered statistically significant ("***", "**", "*", "ns" are "$p < 0.001$" "$p < 0.01$" "$p <$

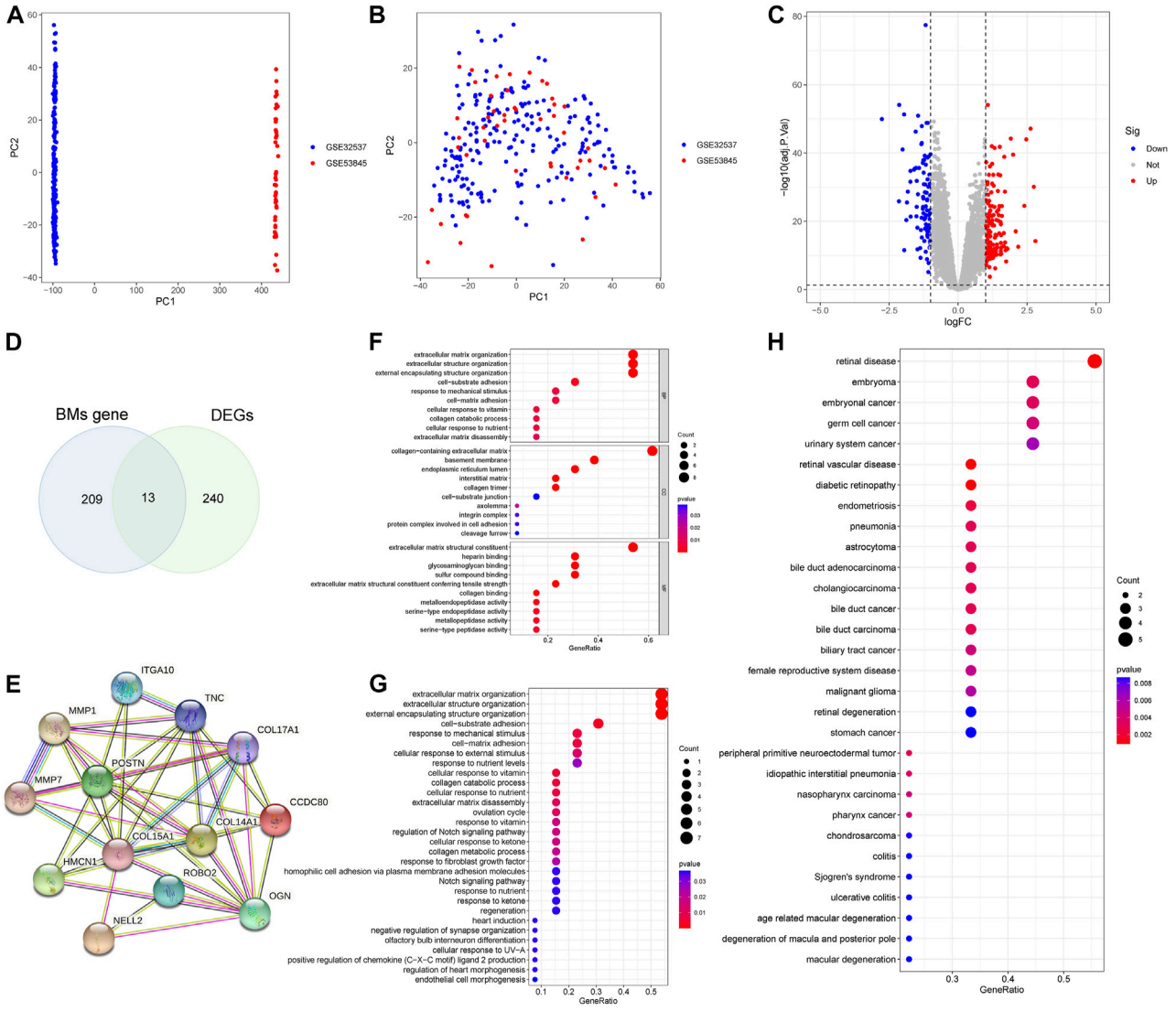


FIGURE 2
Differential expression analysis and enrichment analysis. **(A)** PCA plot shows training set before batch effect. **(B)** PCA plot shows the training set after batch effect. **(C)** A volcanic map of DEGs. **(D)** The Venn diagram of BMs related genes and DEGs. **(E)** Protein interaction network of BMDEGs. GO function**(F)**, KEGG pathway**(G)** and Disease enrichment analysis**(H)** of BMDEGs.

0.05" "no significance"). Relevant scripts and supported data can be seen on the Github website (https://github.com/fuchenkun/Basement-membranes).

3 Results

3.1 Screening for BMDEGs of IPF

We combined GSE32537, GSE53845 datasets and corrected the batch effects for subsequent analyses (Figures 2A, B). 253 DEGs were screened, of which 158 genes were over-expressed and 95 genes were under-expressed. The results of the DEGs were presented as a volcano map (Figure 2C). A Venn diagram was also created, which showed that 13 BMDEGs, of which 12 were upregulated and 1 downregulated (Figure 2D).

3.2 Protein interaction network and enrichment analyses of BMDEGs

We constructed a protein interaction network for BMDEGs (Figure 2E). Then we performed enrichment analyses to better understand the functions, pathways, and diseases that BMDGEs might be involved in. As shown in Figure 2F, our results indicated that biological processes were mainly related to the structure of extracellular matrix, collagen catabolic processes, and responses to mechanical stimulus. The cellular components mainly involved extracellular matrix, basement membrane and endoplasmic reticulum. Extracellular matrix structural constituent, heparin binding, glycosaminoglycan binding, sulfur compound binding and extracellular matrix structural constituent conferring tensile strength were significantly enriched in molecular functions. The KEGG analysis revealed that BMDEGs tended to be enriched in the

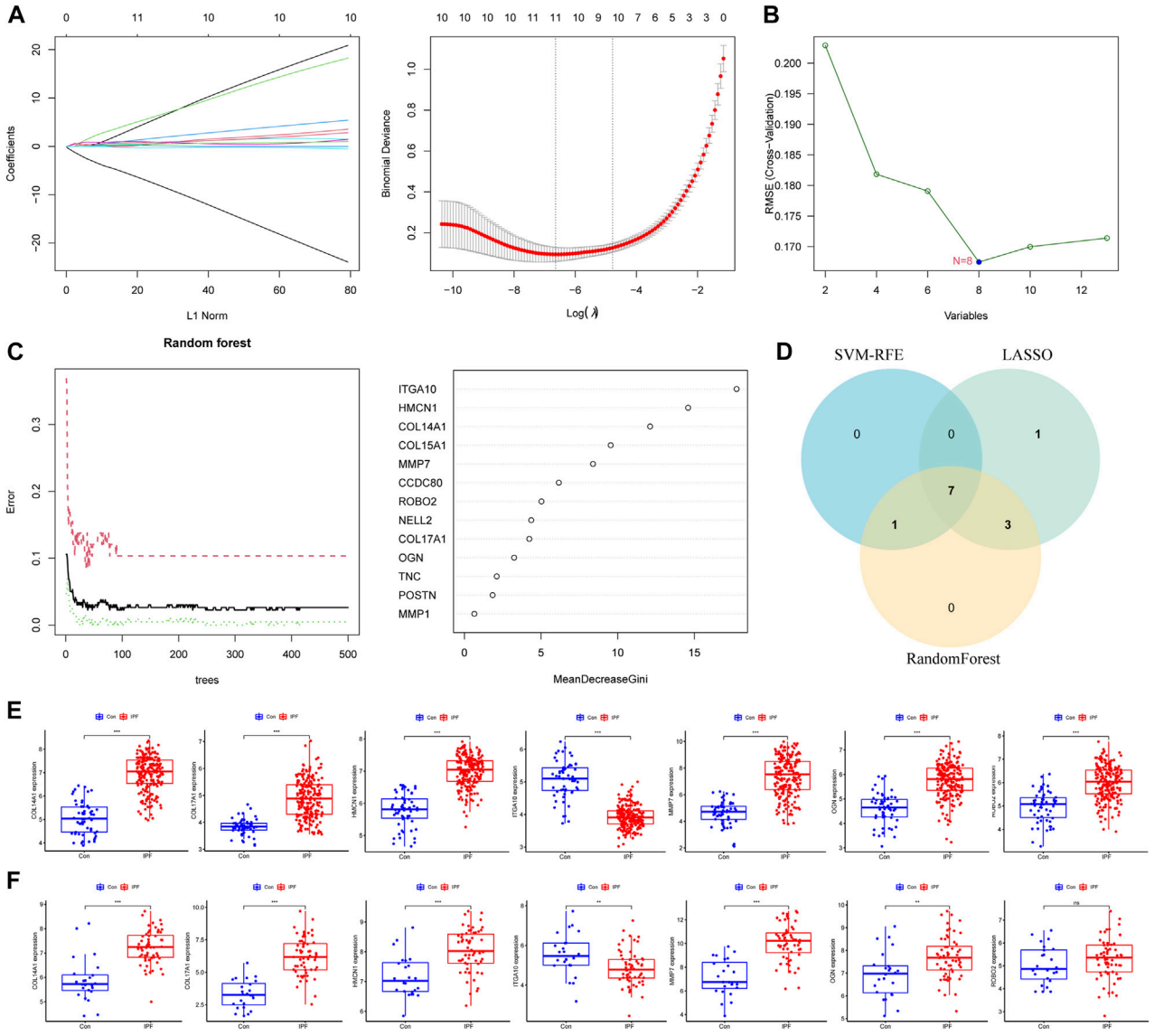


FIGURE 3
Machine learning algorithms identify disease candidate genes. **(A)** LASSO model for screening candidate genes for disease. **(B)** SVM-RFE algorithm for screening disease candidate genes. **(C)** Random forest model to screen candidate genes for disease. **(D)** The Venn diagram shows the common disease candidate genes of LASSO, RandomForest and SVM-RFE algorithms. **(E)** Boxplot representing expression of disease candidate genes in training set. **(F)** Boxplot representing expression of disease candidate genes in the validation set.

following terms: extracellular matrix organization, extracellular structure organization, external encapsulating structure organization, cell–substrate adhesion and response to mechanical stimulus (Figure 2G). Moreover, DO analysis found that BMDEGs were specifically enriched in IPF and were also associated with endocrine disorders, reproductive system diseases, and cancer (Figure 2H).

3.3 Screening and validation for IPF diagnostic markers

We further used machine learning algorithms to screen disease candidate genes from BMDEGs. LASSO regression analysis selected 11 genes (Figure 3A), the SVM-RFE algorithm identified eight genes (Figure 3B), and RandomForest screened 11 genes (Figure 3C). Finally, through the gene intersection obtained by the three algorithms, seven disease candidate genes (*COL14A1*, *COL17A1*, *HMCN1*, *ITGA10*, *MMP7*, *OGN* and *ROBO2*) were identified (Figure 3D). Then we analyzed the expression of seven disease candidate genes in the training group and validated them using external datasets. Validation group dataset eliminates batch effect for subsequent analysis (Supplementary Figure S1). As presented in Figure 3E, the boxplot showed that seven disease candidate genes were significantly upregulated in IPF groups and 1 candidate gene was significantly downregulated in IPF groups. We saw the same results in the validation dataset, but *ROBO2* was not statistically significant (Figure 3F). We further performed ROC analysis to

10.3389/fgene.2023.1114601

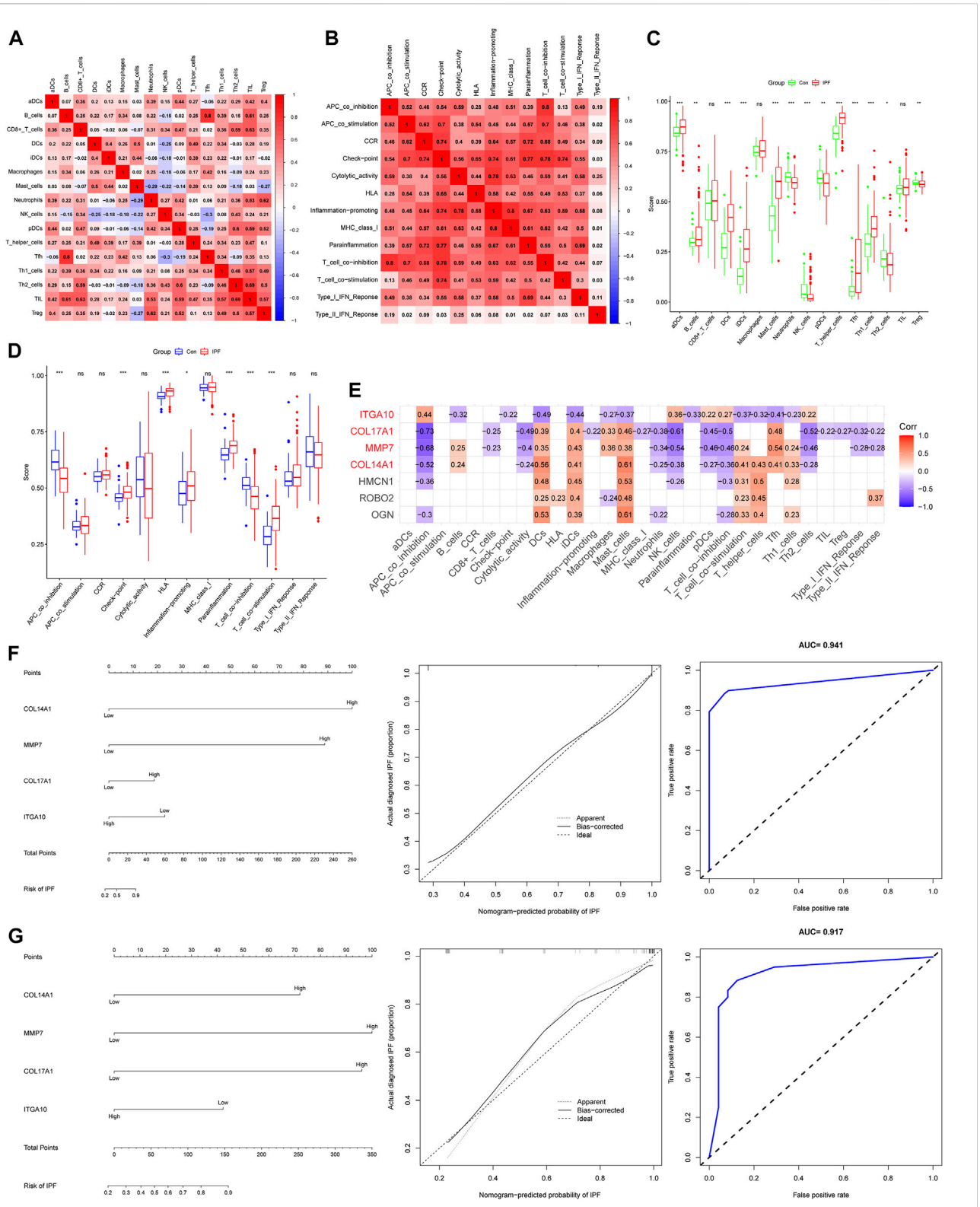


FIGURE 4
Immune infiltration analysis and establishment of IPF risk model. **(A)** Correlation heatmap of immune cells. **(B)** Immune function related heatmap. **(C)** Difference of immune cell expression between IPF and control groups. **(D)** Difference of immune function expression between IPF and control group. **(E)** Heatmap of correlation between IHGs expression and immune infiltration. **(F)** IHGs predicts the occurrence of IPF in training set. **(G)** IHGs predicts the occurrence of IPF the validation set.

examine the diagnostic efficacy of seven disease candidate genes for IPF. The results suggest that the seven disease candidate genes have diagnostic value in distinguishing IPF groups from control groups: *COL14A1* (AUC = 0.964), *COL17A1* (AUC = 0.915), *HMCN1* (AUC = 0.961), *ITGA10* (AUC = 0.946), *MMP7* (AUC = 0.937), *OGN* (AUC = 0.894) and *ROBO2*(AUC = 0.856) (Supplementary Table S1). Similarly, we evaluated the diagnostic efficacy of seven disease candidate genes for IPF in the validation group dataset using ROC analysis. The results indicated that AUCs of the disease candidate genes were *COL14A1* (AUC = 0.881), *COL17A1* (AUC = 0.949), *HMCN1*(AUC = 0.813), *ITGA10* (AUC = 0.707), *MMP7* (AUC = 0.910), *OGN* (AUC = 0.719) and *ROBO2* (AUC = 0.600) (Supplementary Table S2). In conclusion, the AUC values of *COL14A1*, *COL17A1*, *HMCN1*, *ITGA10*, *OGN* and *MMP7* in the training dataset and validation dataset were all greater than 0.7. These results suggest that the candidate genes are closely related to IPF and have the potential to be used as biomarkers of IPF and indicators to evaluate the efficacy of patients.

3.4 Immune infiltration analysis and IHGs screening

We used the ssGSEA algorithm to evaluate immune infiltration in 265 samples (Supplementary Figure S2). In the correlation analysis of immune cells, the positive correlation between Tfh cells and B-cell was the strongest, and the correlation coefficient was 0.8. The negative correlation between Tfh cells and NK cells was the strongest, and the correlation coefficient was −0.3 (Figure 4A). Interestingly, we did not observe a negative correlation for immune function, whereas there was a positive correlation (r = 0.8) between T-cell co inhibition and APC co inhibition (Figure 4B). For immune cells, the expression of aDCs, B-cell, DCs, iDCs, Mast cells, T helper cells, Tfh and Th1 cells were increased in IPF, while the expression of neutrophils, NK cells, pDCs, Th2 cells and Treg cells were decreased (Figure 4C). For immune function, Check point, HLA, Inflammation promoting, T-cell co stimulation and Parainflamation were over-expressed in IPF, while T-cell co inhibition and APC co inhibition were underexpressed in IPF (Figure 4D). Finally, four IHGs were screened by correlation analysis, including *COL14A1*, *COL17A1*, *ITGA10* and *MMP7* (Figure 4E). These results suggest that the activation of multiple immune cells and the coordination of immune functions are important in the pathogenesis of IPF.

3.5 Construction and validation of IPF risk model

We created a logistic regression model of IHGs and used a nomogram to predict IPF risk (Figure 4F). The calibration curve used to evaluate the risk nomogram of IPF patients showed good consistency in this study. The results showed that the AUC of the training data set was 0.941, indicating that our model had good predictive ability. In order to further verify the prediction effect of our model, we used independent external validation dataset to verify (Figure 4G). The results showed that the calibration curve also showed good consistency in the training dataset. The AUC value in the validation dataset was 0.917, which also shows that our model had good predictive ability. Moreover, The C index also shows that our model had good predictive power. The C index was 0.941 (95% CI: 0.917–0.965) in the training dataset and 0.917 (95% CI: 0.840–0.994) in the validation dataset.

3.6 The expression level of IHGs in pan-cancer

To determine whether there are differences in the expression of IHGs in tumors, mRNA expression levels of IHGs in normal and tumor tissues were analyzed. As shown in Figure 5A, the expression of *MMP7* in IHGs was relatively high, while the expression of *ITGA10* was the lowest. The expression of IHGs in different cancer types is also quite different (Figure 5B). Overall, *COL14A1*, *COL17A1*, and *ITGA10* tended to be downregulated in most tumors, while *MMP7* tended to be upregulated in most tumors. Correlation analysis showed that *ITGA10* was weakly positively correlated with *COL14A1*, and weakly negatively correlated with *MMP7* and *COL17A1* (Figure 5C). Although there is a correlation between IHGs, the correlation coefficient value is between −0.15 and 0.23, which proves that the correlation is weak or negligible. *COL14A1* was highly expressed in 1 tumor and lowly expressed in 15 tumors (Figure 5D). *COL17A1* was significantly upregulated in seven tumors, while significantly downregulated in eight tumors (Figure 5E). *ITGA10* expression was increased in six tumors and decreased in eight tumors (Figure 5F). *MMP7* expression was higher in 12 tumors and lower in four tumors (Figure 5G).

3.7 Pan-cancer survival analysis of IHGs

Based on the results of differential analysis, we used forest maps and survival curves to further understand whether IHGs expression was linked with tumor prognosis (Figure 5H). Cox regression analysis revealed that increased *COL14A1* expression was a negative factor affecting KIRP, LGG, BLCA, STAD and OV, while a positive factor affecting ACC. Increased expression of *COL14A1* was related to shorter overall survival (OS) in BLCA, KIRP, LGG and UVM, whereas decreased expression of *COL14A1* was related to shorter OS in ACC and LAML (Figures 6A–F). As shown as Figures 6G–I, Cox regression analysis revealed that *COL17A1* over-expression was an adverse factor for PAAD and SKCM, but a favourable factor for BRCA. Overexpression of *COL17A1* was linked with poorer OS in PAAD and SKCM, whereas increased *COL17A1* expression predicted favorable OS in LGG. As seen as Figures 6J–R, Cox regression analysis found that the increased expression of *ITGA10* was a negative factor for LGG, SARC and KICH, and a positive factor for SKCM and BRCA. High *ITGA10* expression was related to shorter OS for KIRP, LGG, MESO, OV, SARC, STAD, THCA, and longer OS for BRCA and SKCM. According to Figures 6S–Y, Cox regression analysis revealed that high expression of *MMP7* was an adverse factor for PAAD, ACC, LAML, KIRC, LIHC and SKCM, while high expression of *MMP7* was a beneficial factor for BRCA. Survival analysis showed that patients with increased *MMP7* expression in ACC, KIRC, LAML, LGG, LIHC, MESO, PAAD had shorter OS.

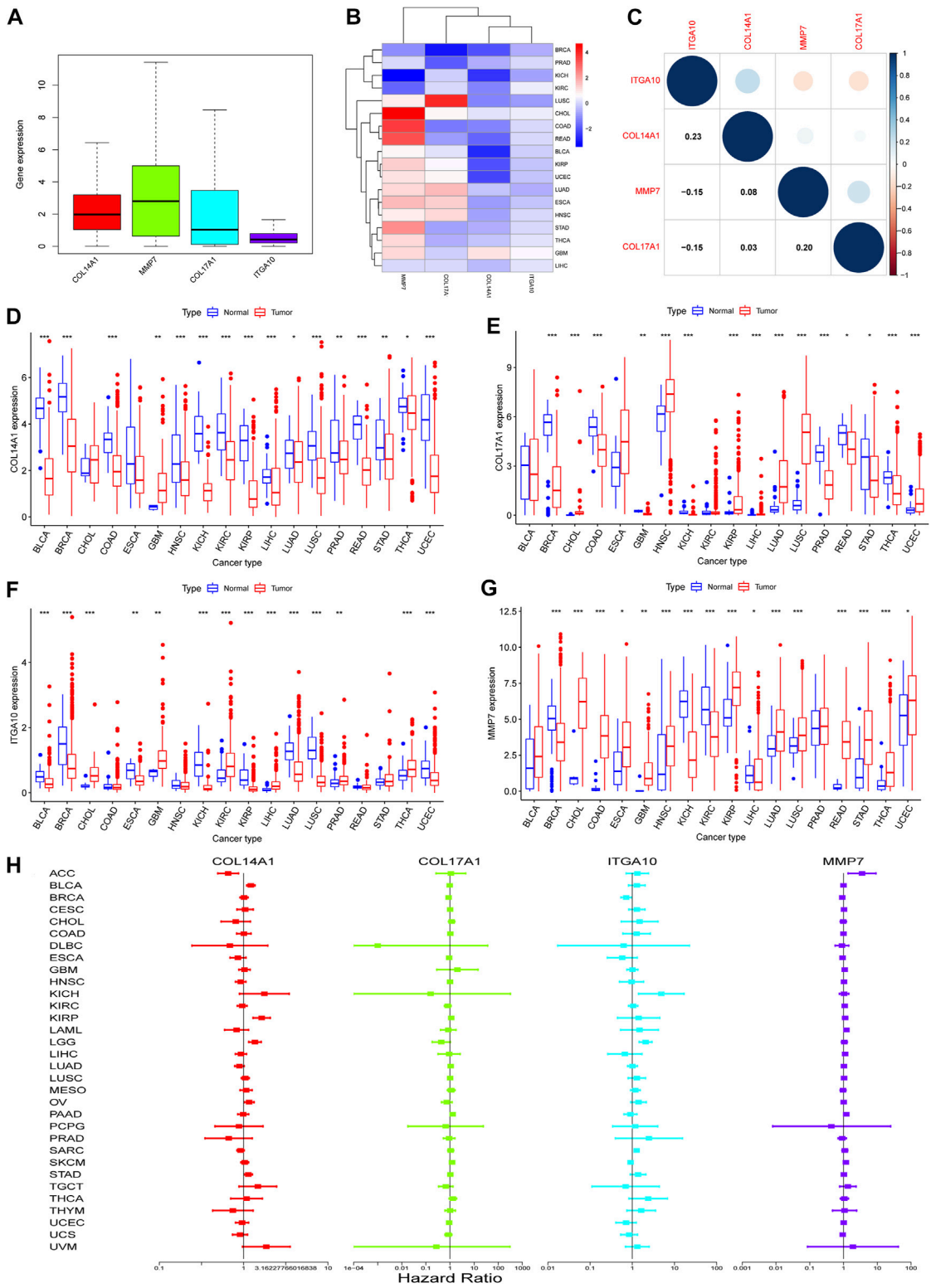


FIGURE 5
Expression of IHGs in Human Cancer. **(A)** Boxplots of IHGs expression levels in cancer. **(B)** Heatmap of IHGs expression levels in different cancer types and adjacent tissues. **(C)** Positive (blue) and negative (red) correlations between IHGs. Expression of *COL14A1***(D)**, *COL17A1***(E)**, *ITGA10***(F)** and *MMP7***(G)** in different tumor types and adjacent tissues. **(H)** Forest plot shows the relationship between IHGs expression and OS in 33 tumors.

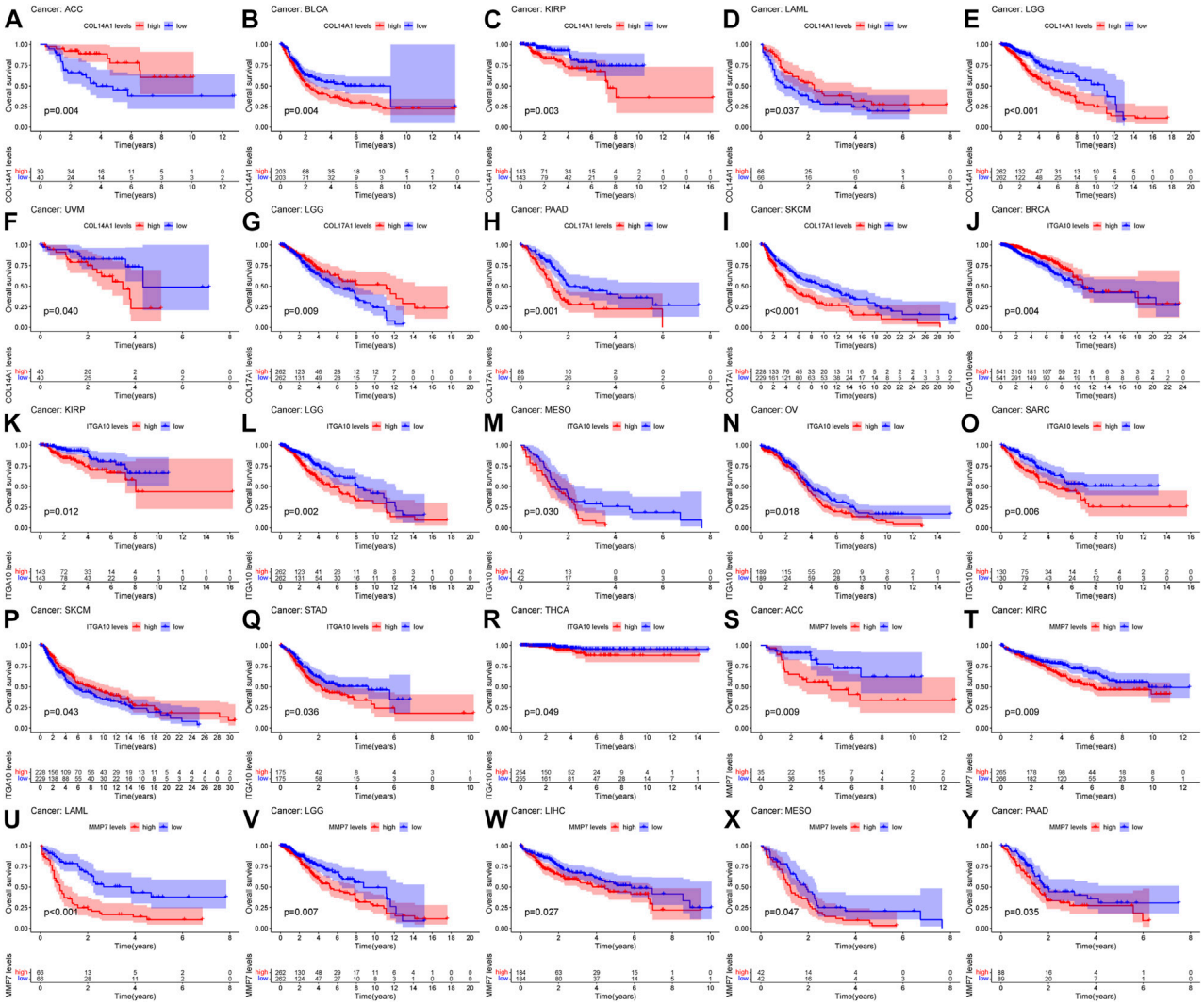


FIGURE 6
Relationship between IHGs expression and prognosis of different tumors. OS survival curves for *COL14A1* in six tumors: **(A)** ACC, **(B)** BLCA, **(C)** KIRP, **(D)** LAML, **(E)** LGG, **(F)** UVM. OS survival curves for *COL17A1* in 3 tumors:**(G)** LGG, **(H)** PAAD, **(I)** SKCM. OS survival curves for *ITGA10* in 9 tumors: **(J)** BRCA, **(K)** KIRP, **(L)** LGG, **(M)** MESO, **(N)** OV, **(O)** SARC, **(P)** SKCM, **(Q)** STAD,**(R)** THCA. OS survival curves for *MMP7* in 7 tumors: **(S)** ACC, **(T)** KIRC, **(U)** LAML, **(V)** LGG, **(W)** LIHC, **(X)** MESO, **(Y)** PAAD.

3.8 Correlation of IHGs expression with TME and tumor stemness

TME is closely related to tumorigenesis and tumor cells escaping the immune system. The therapeutic effect and clinical prognosis of tumor are also correlated with TME. Therefore, the correlation between IHGs expression and tumor purity was assessed to understand whether IHGs are involved in tumor immunity. Overall, IHGs expression was positively related to stromal scores, with *COL14A1* having the strongest correlation with stromal score (Figure 7A). In terms of immune score, IHGs expression was positively correlated with CHOL, KICH, LIHC, and PCPG, etc., (Figure 7B). This suggests that IHGs have similar effects in the TME. Furthermore, we also evaluated the correlation between IHGs expression and tumor stemness score to understand the effect of IHGs expression on tumor differentiation. We found that IHGs were negatively related to RNAss in most tumors (Figure 7C). In contrast, it was positively related to DNAss in KIRC, KIRP, THYM and UVM (Figure 7D). These results suggest that the higher the expression of IHGs, the weaker the stemness score and the higher the degree of tumor differentiation.

3.9 Correlation of IHGs expression with immune subtype and immune checkpoints

Our results showed that *COL14A1*, *COL17A1*, and *MMP7* were over-expressed in the C1, C2, and C6 subtypes, and *ITGA10* was over-expressed in the C5 and C6 subtypes (Figure 7E). The high expression of *COL14A1*, *COL17A1* and *MMP7* was closely related to C1, C2 and C6 subtypes, indicating that these three genes may have carcinogenic effects. Moreover, we observed a significant correlation between IHGs expression and immune checkpoint genes in different

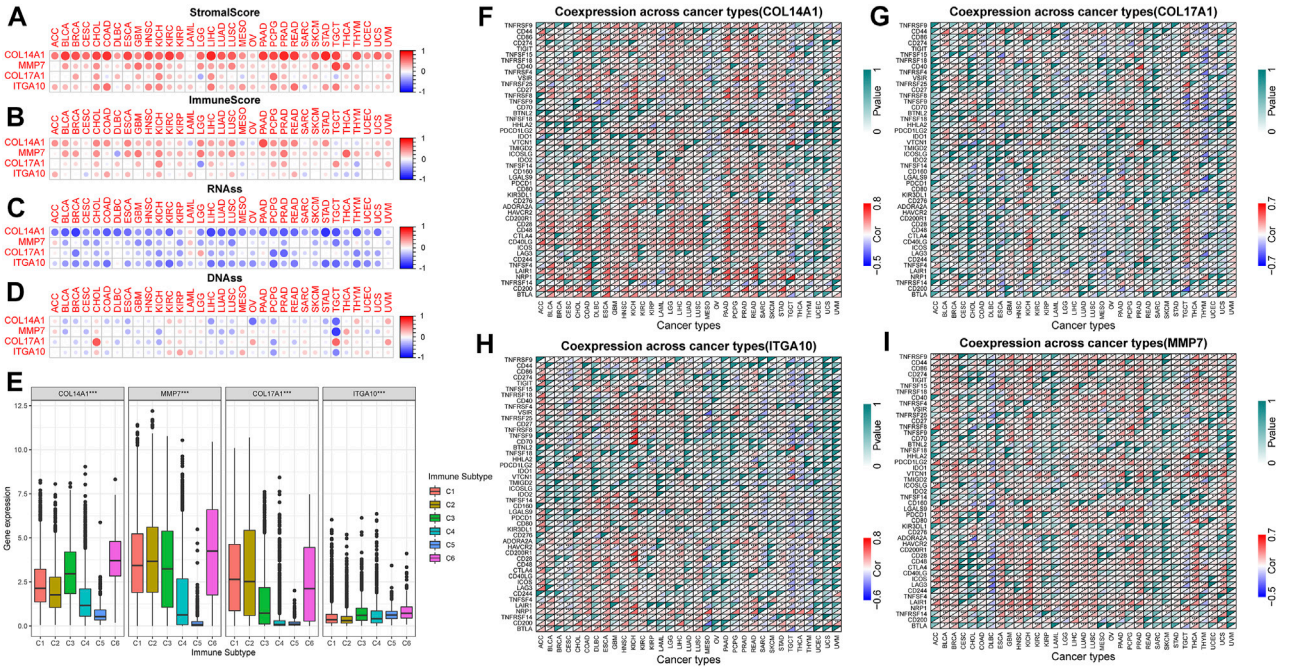


FIGURE 7
Correlation of IHGs expression with TME, stemness score, immune subtypes and immune checkpoints. **(A,B)** IHGs expression was related to immune and stromal scores in 33 tumors. **(C,D)** IHGs expression was related to stemness score in 33 cancers. **(E)** IHGs expression was associated with immune subtypes. The correlation between *COL14A1* **(F)**, *COL17A1* **(G)**, *ITGA10* **(H)** and *MMP7* **(I)** and immune checkpoints in 33 cancers.

tumor types. Specifically, the results revealed that *COL14A1* was positively related to immune checkpoints in most tumors except MESO, OV, SARC, TGCT, THCA, THYM, UCEC, UCS, and UVM (Figure 7F). *COL17A1* was significantly related to immune checkpoints in most tumors except ACC, CHOL, DLBC, UCS, and UVM (Figure 7G). *ITGA10* was positively related to more than 30 immune checkpoint genes in COAD, ESCA, KICH, HNSC and LUSC (Figure 7H). In addition to CESC, CHOL, COAD, SARC, SKCM, and UVM, *MMP7* was closely related to immune checkpoint genes (Figure 7I). These results indicate that IHGs expression is closely related to immune checkpoint genes, suggesting that IHGs may play a vital role in mediating tumor immune patterns.

3.10 Immune cell infiltration analysis of IHGs in pan-cancer

We analyzed the correlation between tumor infiltrating immune cells and IHGs expression by TIMER2.0 to understand whether IHGs participated in tumor immune infiltration. We found that *COL14A1* was positively related to CAF, DCs, Endo, HSC, Macrophage, Monocyte and Tregs (Figure 8A). *COL17A1* was negatively related to most immune cells in HNSC, LUSC and ESCA (Figure 8B). *ITGA10* was positively related to CAF, Endo, HSC, neutrophils and Tregs in most tumors (Figure 8C). *MMP7* was positively related to CAF, DCs, Macrophage and Monocyte in most tumors (Figure 8D). Compared with *COL17A1*, *ITGA10* and *MMP7*, *COL14A1* had a higher correlation coefficient with infiltrating cells. These results suggest a potential mechanism by which IHGs have different prognostic value in different tumors.

3.11 TMB and MSI analysis of IHGs in human cancers

The more frequent the mutation of tumor cells, the more new antigens produced, making them more susceptible to immunotherapy (Gryfe et al., 2000; Samstein et al., 2019). Herein, we investigated the correlation between IHGs expression with TMB, MSI. We found that *COL14A1*/*ITGA10*/*MMP7* was positively related to TMB in LGG, and negatively related to TMB in STAD, LUSC, LUAD, LIHC, HNSC and BLCA (Figures 8E–H). As for MSI, *COL14A1*/*ITGA10*/*MMP7* expression was negatively correlated with STAD (Figures 8I–L). Although these correlations are important to guide immunotherapy in cancer patients, the correlation coefficients between IHGs expression and TMB and MSI did not exceed 0.6 in all tumor types, suggesting that IHGs are unlikely to affect tumorigenesis by participating in gene modification processes and are insufficient to independently predict patient response to immunotherapy.

3.12 Functional enrichment analysis of IHGs in pan-cancer

We utilized GeneMANIA to screen genes associated with IHGs for comprehensive functional and pathway analysis of IHGs. Finally,

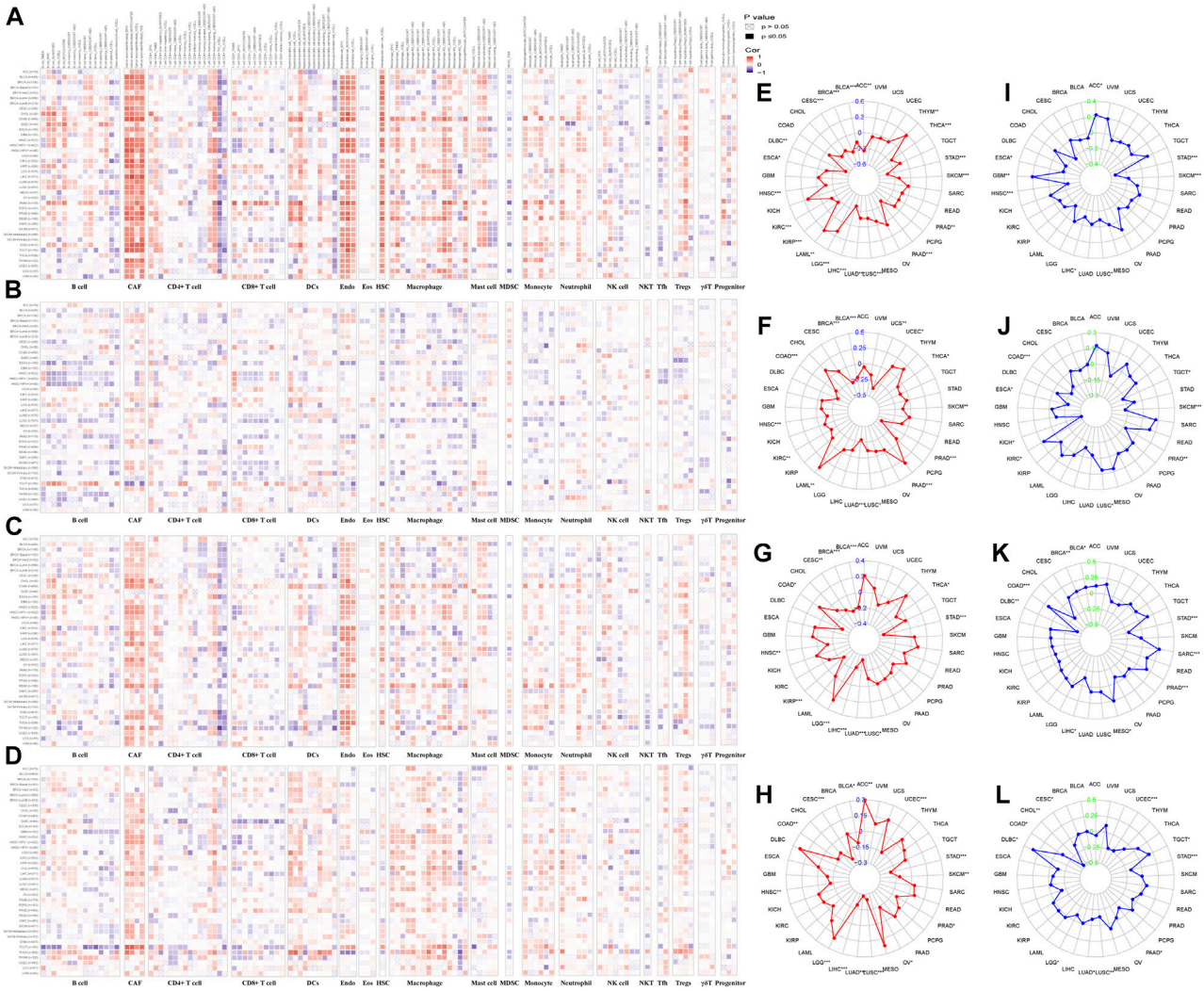


FIGURE 8
Correlation of IHGs expression with immune cell infiltration, TMB and MSI. The correlation between *COL14A1* **(A)**, *COL17A1* **(B)**, *ITGA10* **(C)** and *MMP7* **(D)** and 21 immune cells in 33 cancers. The correlation between *COL14A1* **(E)**, *COL17A1* **(F)**, *ITGA10* **(G)** and *MMP7* **(H)** and TMB in 33 cancers. The correlation between *COL14A1* **(I)**, *COL17A1* **(J)**, *ITGA10* **(K)** and *MMP7* **(L)** and MSI in 33 cancers.

we constructed an IHGs-centric PPI network consisting of 24 genes (Figure 9A). Metascape significantly enriched items include Extracellular matrix organization, ECM-receptor interaction, Degradation of the extracellular matrix, Integrin cell surface intercations, epidermis development, PID AJDISS 2PATHWAY, ECM proteoglycans, response to wounding, Proteoglycans in cancer, Wnt signaling pathway and pluripotency, Type I hemidesmosome assembly, Immunoregulatory interactions between a Lymphoid, appendage development and response to growth factor (Figure 9B).

3.13 Drug response analysis of IHGs

We further explored whether the expression of IHGs has guiding significance for clinical medication. Among 263 drugs (FDA approved or clinical trials), we found that the sensitivity of 55 drugs was significantly correlated with IHGs expression levels. As presented in Figure 9C, we show the top 20 drugs most significantly associated with IHGs. The upregulated expression of IHGs was associated with increased sensitivity to Afatinib, Pentostatin, Gefitinib, kahalide f, Lapatinib, Vandetanib, Ibrutinib, Erlotinib, and okadaic acid, and decreased sensitivity to Pyrazoloacridine, AT−13387, Pazopanib, Vincristine, Lapachone, Etoposide, Lomustine, Paclitaxel, Epothilone B, and Pipamperone. Overall, we found that IHGs expression was associated with treatment response, suggesting that IHGs may be involved in tumor drug resistance.

4 Discussion

In this study, we screened 13 BMDEGs using bioinformatics methods, among which 12 genes were over-expressed and 1 gene was lowly expressed. Subsequent GO enrichment analysis showed

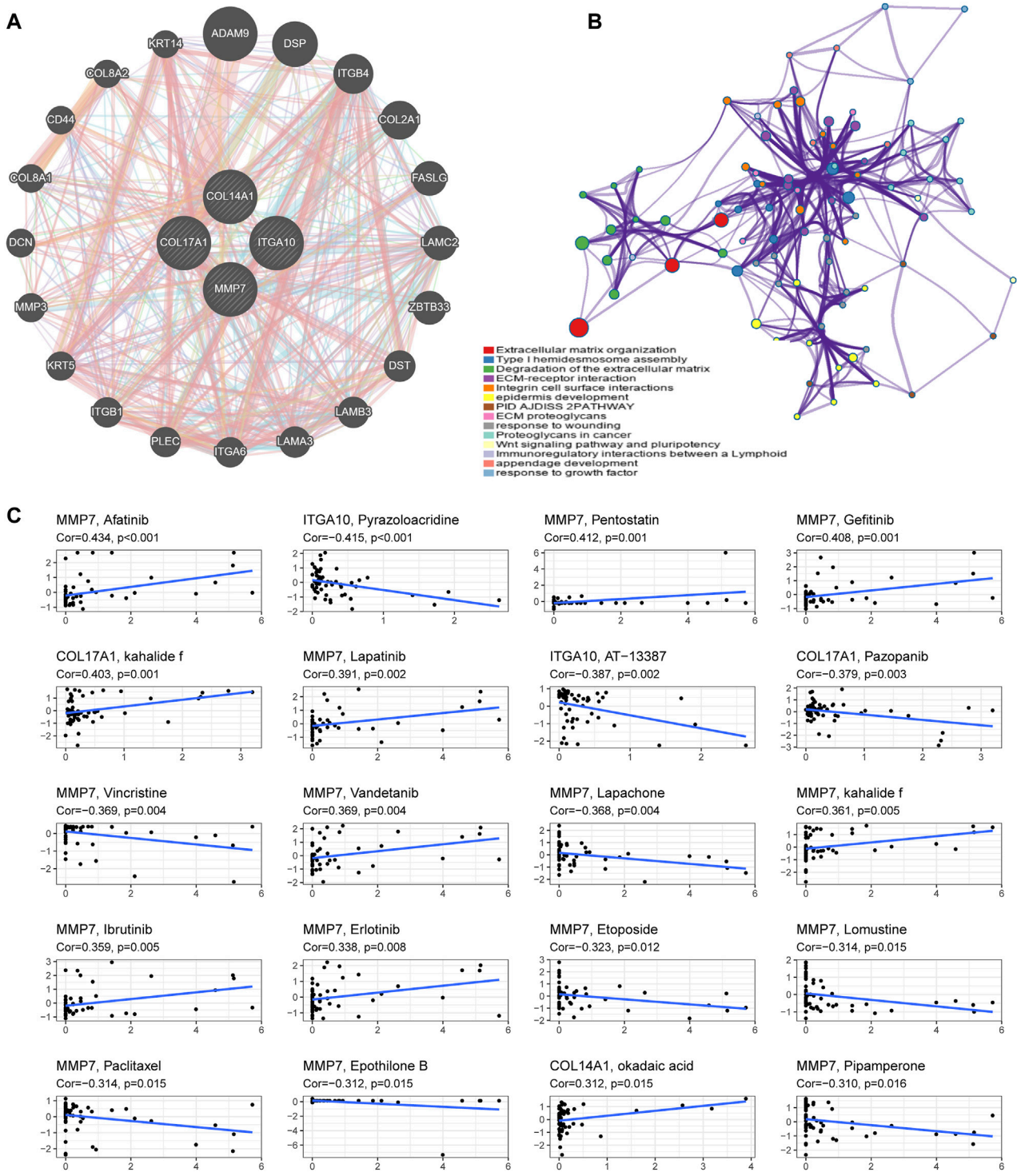


FIGURE 9
Functional enrichment analysis and drug sensitivity analysis of IHGs. **(A)** Construction of a PPI network with 24 genes centered on IHGs. **(B)** Enrichment analysis using Matescape. **(C)** Correlation analysis between IHGs and drug sensitivity of anticancer drugs in CellMiner.

that all BMDEGs were mainly related to extracellular matrix tissue and collagen catabolic process, while KEGG enrichment analysis showed a certain correlation with Notch signaling pathway. Based on three machine learning algorithms, we screened seven disease candidate genes. Using external datasets, it was confirmed that *COL14A1*, *COL17A1*, *HMCN1*, *MMP7*, *OGN* and *ROBO2* were highly expressed in IPF, while *ITGA10* was lowly expressed in IPF. ROC curve analysis further confirmed that all disease candidate genes have diagnostic value in IPF, suggesting that they may have potential application prospects in the treatment of

IPF. The findings suggest their potential usage for diagnostic value in IPF since they are highly expressed in IPF. The findings will eventually lead to future studies about the potential role of their involvement in IPF, and if their roles are confirmed, the potential application prospects in the treatment of IPF are likely to evaluated. Finally, four IHGs (*COL14A1*, *COL17A1*, *ITGA10*, *MMP7*) were screened out. We then construct a logistic regression model of IHGs and use a nomogram to predict IPF risk. The AUC of the training set was 0.941, and that of the verification set was 0.917. It shows that our model has good predictive ability, and these four genes are potential biomarkers of IPF. Our study provides a theoretical basis for studying the role of BMs-associated immune biomarkers in the pathogenesis of IPF, and provides promising research suggestions for subsequent studies.

In pan-cancer, the expression of IHGs was significantly different across cancers compared to comparable normal tissues. *COL14A1* was significantly downregulated in almost all cancer types. *COL17A1*, *ITGA10* and *MMP7* showed high intertumoral heterogeneity between tumor tissues and adjacent tissues. These results suggest that IHGs are a potential cancer biomarker.

Our study found that *COL14A1* was an adverse factor for KIRP, LGG, BLCA, STAD and OV, and a favourable factor for ACC. However, there is a lack of relevant studies to support the effect of *COL14A1* on the prognosis of these cancers, and more studies are needed to prove this. *COL14A1* encodes the alpha chain of type XIV collagen, which is linked with mature collagen fibers (Schuppan et al., 1990). It has been reported that *COL14A1* can affect arterial remodeling and participate in the occurrence of cardiovascular diseases (Weis-Müller et al., 2006; Guay et al., 2015). To our knowledge, there is a lack of literature on the role of *COL14A1* in IPF. When compared cancer with or without metastasis, it seems that further decrease of *COL14A1* has better outcome, but, this has to be tested in a larger scale to validate (Goto et al., 2015; Jiang et al., 2022). *COL14A1* methylation is an unfavorable prognostic factor for renal cell carcinoma, and low *COL14A1* expression seems to promote tumorigenesis of renal cell carcinoma (Morris et al., 2010). It seems that *COL14A1* is increased in IPF, while it is decreased in cancer. The differential expression of *COL14A1* may indicate the critical signaling that differentiates IPF - a disease with non-stopping fibroblast growth, from cancer-a disease with non-stopping malignant cell growth. Therefore, our studies suggest the critical and differentiating signaling may involve *COL14A1*, its function or signaling, and hopefully, that can be investigated by future studies.

COL17A1 is one of the triple-helix collagen genes encoding collagen XVII, a type II transmembrane protein found in basal epithelial cells that can affect cell growth and migration (Natsuga et al., 2019; Kozawa et al., 2021). *COL17A1* also lacks relevant research in the field of IPF. Currently, research on *COL17A1* has focused on cancer and skin diseases (Nishie, 2020). Our study found that *COL17A1* over-expression was related to poor prognosis of SKCM and PAAD. Studies have shown that *COL17A1* is over-expressed in a variety of cancers (Thangavelu et al., 2016; Huang et al., 2022). In contrast, another study found that upregulation of *COL17A1* expression was related to better prognosis in breast cancer (Yodsurang et al., 2017). *COL17A1* inhibits cancer cell migration and invasion by inactivating AKT/mTOR pathway, and its over-expression is linked with longer survival in patients with invasive breast cancer (Lothong et al., 2021).

ITGA10 is a type II collagen-binding integrin first isolated from chondrocytes (Camper et al., 1998). *ITGA10* has the highest content in cartilage tissue and plays a crucial part in the formation of growth plates during bone development (Bengtsson et al., 2005). From the current overall research situation, the research on *ITGA10* mainly focuses on cancer. Our study suggested that increased *ITGA10* expression was linked with poor prognosis of SARC and longer survival of BRCA and SKCM. Previous studies have reported that *ITGA10* promotes drug resistance and proliferation of osteosarcoma cells by the activation of PI3K/AKT signaling pathway (Li et al., 2021). Similarly, *ITGA10* promotes myxfibrosarcoma survival and metastasis by activating TRIO/RAC and RICTOR signaling pathways, and antitumor effects were observed in mouse xenografts after *ITGA10* inhibition (Okada et al., 2016). In addition, dysregulation and carcinogenic effects of *ITGA10* had been observed in lung cancer, prostate cancer, and thyroid cancer (Mertens-Walker et al., 2015; Su et al., 2015; Saftencu et al., 2019). These findings suggest that *ITGA10* acts as an oncogene. According to our knowledge, *ITGA10* has not been reported on IPF, and more attention should be paid to IPF.

Different from *COL14A1*, *COL17A1* and *ITGA10*, *MMP7* had been supported by numerous literatures in the research field of IPF. *MMP7* is the smallest member of the matrix metalloproteinase family. *MMP7* plays a crucial part in the pathogenesis of fibrosis by degrading extracellular matrix proteins and activating multiple signaling molecules (Niu et al., 2019; Mahalanobish et al., 2020). *MMP7* is a target gene of Wnt/β-catenin and highly expressed in proliferative epithelial cells of IPF (Zuo et al., 2002; Fujishima et al., 2010). Previous studies have identified *MMP7* as a potential biomarker for IPF. For instance, *MMP-7* was identified as a predictor of survival in a combined model incorporating clinical parameters and *MUC5B* genotype (Peljto et al., 2013; Biondini et al., 2021). Consistent with previous studies, we found increased *MMP7* expression in IPF and negatively correlated with decreased FVC (Bauer et al., 2017). Some studies have shown that *MMP7* combined with other biomarkers may improve the survival prediction of IPF patients (Song et al., 2013; Hamai et al., 2016). In addition, a phase II clinical study showed that *MMP7* protein levels decreased in a dose-dependent manner after using JNK inhibitors, indicating that the use of *MMP7* to track IPF progression has potential clinical benefits (van der Velden et al., 2016). However, a recent study found that there was no difference in the baseline concentration of *MMP7* between IPF patients with or without disease progression, and short-term changes in its concentration could not reflect disease progression (Raghu et al., 2018; Khan et al., 2022). Another study indicated that *MMP7* was over-expressed in patients with subclinical interstitial lung disease and under-expressed in patients with mature IPF compared to healthy controls (Drakopanagiotakis et al., 2018). It is suggested that *MMP7* can be used as a potential marker for early detection of IPF. Besides, more and more evidences support *MMP7* as an oncogene involved in tumor cell proliferation, migration and apoptosis (Sun et al., 2021; Van Doren, 2022). *MMP7* is highly expressed in many cancers, and its expression is related to survival time and tumor stage (Lee et al., 2006; Liao et al., 2021). Knockdown of *MMP7* gene can inhibit tumor proliferation, migration and reduce drug resistance (Sanli et al., 2013; Yuan et al., 2020). Therefore, *MMP7* is expected to become a potential biomarker for evaluating tumor prognosis and a new target for tumor therapy.

GO enrichment analysis showed that BMDEGs were involved in the composition of basement membrane, extracellular matrix and endoplasmic reticulum cavity, and were related to extracellular matrix tissue, collagen metabolism and metallopeptidase activity. Alveolar epithelial damage and abnormal tissue repair are considered to be key factors in the development of IPF, which ultimately leads to the recruitment and activation of myofibroblasts to produce collagen-rich extracellular matrix. The deposition of extracellular matrix in IPF mainly involves matrix metalloproteinases (MMPs) and tissue inhibitors of metalloproteinases (TIMPs). More and more studies support the key role of MMPs in the pathogenesis of pulmonary fibrosis. Interestingly, some MMPs have pro-fibrotic effects, while others seem to play a protective role. In our study, we found that *MMP7* was over-expressed in IPF patients, and previous studies have shown that it contributes to the progression and adverse consequences of IPF, while *MMP19* seems to have a protective effect (Jara et al., 2015). Changes in the extracellular matrix of IPF can also affect the transcription of lung fibroblasts, resulting in abnormal translation of ECM proteins (Zolak and de Andrade, 2012). In addition, the degradation product of ECM (matrikines) also acts as a signal molecule and plays a central role in the fibrosis of IPF. Existing evidence suggests that ECM plays an important role in driving the circulation of pathogenic disease signals mediated by integrins, growth factors, matrikines, and MMPs (Hewlett et al., 2018). It is worth noting that the increased stiffness of ECM tissue is also a key driver of the fibrosis process. Compared with healthy lung scaffolds, collagen, proteoglycan and ECM glycoprotein in IPF scaffolds increased, but specific BMs proteins (such as laminin and collagen IV) decreased (Elowsson Rendin et al., 2019). New treatments for these ECM-driven processes are expected to bring benefits to IPF patients. Moreover, KEGG enrichment analysis showed that BMDEGs were related to Notch signaling pathway. It is reported that Notch signaling pathway plays a key role in the development, balance and regeneration of the respiratory system (Kiyokawa and Morimoto, 2020). The disorder of Notch signaling pathway is related to the occurrence of IPF, and the activation of Notch signaling can accelerate pulmonary fibrosis (Yang et al., 2022). Therefore, regulating the activation of Notch signaling pathway may be a new anti-fibrosis treatment strategy.

Although the pathogenesis of IPF remains unclear, a growing number of studies have implicated immune activation in its pathogenesis. Therefore, we used ssGSEA to further dissect the immune infiltration of the disease. In our results, we found that the expression of IHGs was correlated with different degrees of immune cell infiltration. The role of neutrophils in IPF remains unclear. On the one hand, inhibition of neutrophil chemokine *CXCL8* or lack of neutrophil elastase can reduce bleomycin-induced pulmonary fibrosis (Gregory et al., 2015; Gschwandtner et al., 2017). Similarly, increased neutrophils are associated with decreased FVC and all-cause mortality in IPF patients (Achaiah et al., 2022). It has been reported that deletion of exon 18 of *COL17A1* in mice leads to IL-17-related inflammatory responses in the skin and infiltration of eosinophils, neutrophils, T-cell and mast cells (Lindgren et al., 2023). XIV collagen is a neutrophil chemotactic factor that plays a role in neutrophil recruitment in rat inflammation (Nakagawa et al., 1999). On the other hand, the increase of neutrophils in BALF was not significantly related to the survival rate of IPF patients (Tabuena et al., 2005). Fibrosis caused mice infected with bacteria to show a higher mortality rate through the destruction of neutrophils (Warheit-Niemi et al., 2022). However, whether neutrophils have prognostic value in IPF is unclear, and further studies are needed to confirm the role of neutrophils in IPF. B-cell were increased in patients with IPF, which is consistent with our current findings. Activation of immune responses and increased infiltration of B-cell and macrophages are associated with IPF development (Xu et al., 2021). Previous studies have shown that B-cell and BLyS are elevated in patients with IPF and are inversely associated with patient outcome (Heukels et al., 2019). Consistent with increased B-cell activation, plasma IgA was elevated in IPF patients and inversely correlated with FVC (Heukels et al., 2019). These findings suggest that inhibition of B-cell activation has potential therapeutic value for IPF. In addition, our study showed a decrease in Tregs in IPF patients, which is consistent with previous findings. There is conflicting evidence supporting the role of Tregs in IPF. Tregs were originally thought to have anti-fibrotic effects. Tregs in peripheral blood and BALF were significantly reduced in IPF patients and were associated with decreased FVC (Kotsianidis et al., 2009). Subsequent studies have shown that tregs can promote fibrosis. The decrease of Tregs in peripheral blood and BALF of IPF patients is related to the degree of fibrosis (Peng et al., 2014) In the bleomycin-induced PF model, depletion of Tregs resulted in reduction of fibrosis, while induction or metastasis of Tregs resulted in worsening of fibrosis (Birjandi et al., 2016). Tregs may play different roles in different stages of fibrosis. Tregs play a pro-fibrotic role in the early stage of PF and a protective role in the late stage (Boveda-Ruiz et al., 2013). Tregs can promote collagen deposition and release of TGF-β in the early stage of PF (Lo Re et al., 2011). Tregs depletion attenuates PF by promoting Th17 response and regulating the shift of disturbed Th1/Th2 balance to Th1 dominance in lung tissue (Xiong et al., 2015). Consequently, we speculate that Tregs regulate different T-cell subsets at different stages of pulmonary fibrosis, which may explain the different roles of Tregs in pulmonary fibrosis. Existing studies have shown that inhibition of macrophage migration inhibitory factor (MIF) can downregulate the expression level of *ITGA10* and reduce bleomycin-induced pulmonary inflammation and fibrosis in rats (Luo et al., 2021). The relationship between *ITGA10* and MIF is still lacking relevant evidence and the future development of inhibitors targeting MIF may contribute to the treatment of pulmonary fibrosis. It has been reported that activated *MMP7* is located on alveolar macrophages and proliferative epithelial cells (Fujishima et al., 2010). Future studies are needed to further confirm how IHGs participate in the pathological process of IPF by affecting immunity.

We further explored the correlation between IHGs expression and TME, immune subtypes and immune cell infiltration. Our study found that IHGs expression was linked with different levels of immune and stromal cell infiltration. Further analysis revealed that *COL14A1* was positively correlated with CAF, DCs, Endo, HSC, Macrophage, Monocyte and Tregs. *ITGA10* was positively correlated with CAF, Endo, HSC, Neutrophil and Tregs. *MMP7* was positively correlated with CAF, DCs, Macrophage and Monocyte. The correlation between *COL17A1* and immune cells was not prominent. In addition, we also found that *COL14A1*, *COL17A1*, and *MMP7* were associated with more invasive immune subtypes, including C1, C2, and C6 subtypes. These results suggest that changes in immune and stromal cell composition

make IHGs have different clinical features and immunotherapy responses. The relationship between IHGs expression and TME needs to be further studied at cellular and molecular levels.

We also explored the relationship between IHGs expression and immune checkpoint, tumor stemness score, TMB, and MSI. The results showed that IHGs expression was significantly related to RNAss and DNAss in most tumors. Previous studies have reported that higher stemness scores are linked with stronger tumor stem cell dedifferentiation and active biological processes, suggesting potential targets for chemotherapy drug development in cancer patients (Malta et al., 2018). IHGs expression level and immune checkpoint analysis showed that there was significant correlation between IHGs expression level and immune checkpoint in different tumors. Previous studies have shown that TMB is a good biomarker for predicting immunotherapy response in tumor patients (Goodman et al., 2017; Chan et al., 2019). MSI is also associated with prognosis, and high MSI indicates better prognosis (Ganesh et al., 2019). Our study suggested that IHGs expression was significantly related to TMB and MSI. These results reveal that immunotherapy may have potential benefits for cancer patients and may also help clinicians quickly identify patients who respond to immunotherapy. However, more studies at the molecular level are needed to fully explain the relationship between IHGs and immune response.

However, our research also has some shortcomings. Firstly, the biological mechanisms of *COL14A1*, *COL17A1*, *ITGA10* and *MMP7* in IPF and cancer are still unclear. Second, the results of this study need to be confirmed by relevant animals and human experiments. In the future research, we will continue to pay attention to the role of *COL14A1*, *COL17A1*, *ITGA10* and *MMP7* in IPF.

5 Conclusion

In summary, our study shows that BMs and immune disorders are closely associated with IPF. The IPF risk model based on IHGs showed that the high expression of *COL14A1*, *COL17A1*, *ITGA10* and *MMP7* was positively related to the risk of IPF. It was further confirmed that the AUC of the training set was 0.941 and that of the verification set was 0.917, indicating that *COL14A1*, *COL17A1*, *ITGA10* and *MMP7* were potential biomarkers for predicting the risk of IPF. Pan-cancer analysis showed that IHGs were related to prognosis, immune infiltration and drug sensitivity of cancer patients, and were expected to become new biomarkers for cancer patients. However, multicenter, large-scale and prospective studies are needed to confirm our results before *COL14A1*, *COL17A1*, *ITGA10* and *MMP7* are applied clinically.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: The row data included in this study are available in GEO (https://www.ncbi.nlm.nih.gov/geo/, (accessed on 16 August 2022)) and UCSC Xena (https://xena.ucsc.edu/, (accessed on 24 June 2022)).

Author contributions

YC designed the study, WY performed the analysis procedures, BC contributed to the revision of this article, HZ and XW contributed analysis tools, LC analyzed the results, and CF contributed to the writing of the manuscript. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

The author would like to thank Pudan for his valuable advice on data analysis.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1114601/full#supplementary-material

References

Achaiah, A., Rathnapala, A., Pereira, A., Bothwell, H., Dwivedi, K., Barker, R., et al. (2022). Neutrophil lymphocyte ratio as an indicator for disease progression in Idiopathic Pulmonary Fibrosis. *BMJ Open Respir. Res.* 9 (1), e001202. doi:10.1136/bmjresp-2022-001202

Allen, R. J., Guillen-Guio, B., Oldham, J. M., Ma, S. F., Dressen, A., Paynton, M. L., et al. (2020). Genome-wide association study of susceptibility to idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* 201 (5), 564–574. doi:10.1164/rccm.201905-1017OC

Ballester, B., Milara, J., and Cortijo, J. (2019). Idiopathic pulmonary fibrosis and lung cancer: Mechanisms and molecular targets. *Int. J. Mol. Sci.* 20 (3), 593. doi:10.3390/ijms20030593

Banerjee, S., Lo, W. C., Majumder, P., Roy, D., Ghorai, M., Shaikh, N. K., et al. (2022). Multiple roles for basement membrane proteins in cancer progression and EMT. *Eur. J. Cell Biol.* 101 (2), 151220. doi:10.1016/j.ejcb.2022.151220

Bargagli, E., Piccioli, C., Rosi, E., Torricelli, E., Turi, L., Piccioli, E., et al. (2019). Pirfenidone and Nintedanib in idiopathic pulmonary fibrosis: Real-life experience in an Italian referral centre. *Pulmonology* 25 (3), 149–153. doi:10.1016/j.pulmoe.2018.06.003

Bauer, Y., White, E. S., de Bernard, S., Cornelisse, P., Leconte, I., Morganti, A., et al. (2017). MMP-7 is a predictive biomarker of disease progression in patients with idiopathic pulmonary fibrosis. *ERJ Open Res.* 3 (1), 00074. doi:10.1183/23120541.00074-2016

Bengtsson, T., Aszodi, A., Nicolae, C., Hunziker, E. B., Lundgren-Akerlund, E., and Fässler, R. (2005). Loss of alpha10beta1 integrin expression leads to moderate dysfunction of growth plate chondrocytes. *J. Cell Sci.* 118 (5), 929–936. doi:10.1242/jcs.01678

Biondini, D., Cocconcelli, E., Bernardinello, N., Lorenzoni, G., Rigobello, C., Lococo, S., et al. (2021). Prognostic role of MUC5B rs35705950 genotype in patients with idiopathic pulmonary fibrosis (IPF) on antifibrotic treatment. *Respir. Res.* 22 (1), 98. doi:10.1186/s12931-021-01694-z

Birjandi, S. Z., Palchevskiy, V., Xue, Y. Y., Nunez, S., Kern, R., Weigt, S. S., et al. (2016). CD4(+)CD25(hi)Foxp3(+) cells exacerbate bleomycin-induced pulmonary fibrosis. *Am. J. Pathol.* 186 (8), 2008–2020. doi:10.1016/j.ajpath.2016.03.020

Boveda-Ruiz, D., D'Alessandro-Gabazza, C. N., Toda, M., Takagi, T., Naito, M., Matsushima, Y., et al. (2013). Differential role of regulatory T cells in early and late stages of pulmonary fibrosis. *Immunobiology* 218 (2), 245–254. doi:10.1016/j.imbio.2012.05.020

Camper, L., Hellman, U., and Lundgren-Akerlund, E. (1998). Isolation, cloning, and sequence analysis of the integrin subunit alph a10, a beta1-associated collagen binding integrin expressed on chondro cytes. *J. Biol. Chem.* 273 (32), 20383–20389. doi:10.1074/jbc.273.32.20383

Chan, T. A., Yarchoan, M., Jaffee, E., Swanton, C., Quezada, S. A., Stenzinger, A., et al. (2019). Development of tumor mutation burden as an immunotherapy biomarker: Utility for the oncology clinic. *Ann. Oncol.* 30 (1), 44–56. doi:10.1093/annonc/mdy495

Chen, H., Qu, J., Huang, X., Kurundkar, A., Zhu, L., Yang, N., et al. (2016). Mechanosensing by the α6-integrin confers an invasive fibroblast phenotype and mediates lung fibrosis. *Nat. Commun.* 7, 12564. doi:10.1038/ncomms12564

Collins, S. L., Chan-Li, Y., Hallowell, R. W., Powell, J. D., and Horton, M. R. (2012). Pulmonary vaccination as a novel treatment for lung fibrosis. *PLoS One* 7 (2), e31299. doi:10.1371/journal.pone.0031299

Drakopanagiotakis, F., Wujak, L., Wygrecka, M., and Markart, P. (2018). Biomarkers in idiopathic pulmonary fibrosis. *Matrix Biol.* 68-69, 404–421. doi:10.1016/j.matbio.2018.01.023

Elowsson Rendin, L., Löfdahl, A., Åhrman, E., Müller, C., Notermans, T., Michaliková, B., et al. (2019). Matrisome properties of scaffolds direct fibroblasts in idiopathic pulmonary fibrosis. *Int. J. Mol. Sci.* 20 (16), 4013. doi:10.3390/ijms20164013

Fujishima, S., Shiomi, T., Yamashita, S., Yogo, Y., Nakano, Y., Inoue, T., et al. (2010). Production and activation of matrix metalloproteinase 7 (matrilysin 1) in the lungs of patients with idiopathic pulmonary fibrosis. *Arch. Pathol. Lab. Med.* 134 (8), 1136–1142. doi:10.5858/2009-0144-OA.1

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 (6), 361–375. doi:10.1038/s41575-019-0126-x

Goodman, A. M., Kato, S., Bazhenova, L., Patel, S. P., Frampton, G. M., Miller, V., et al. (2017). Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. *Mol. Cancer Ther.* 16 (11), 2598–2608. doi:10.1158/1535-7163.MCT-17-0386

Goto, R., Nakamura, Y., Takami, T., Sanke, T., and Tozuka, Z. (2015). Quantitative LC-MS/MS analysis of proteins involved in metastasis of B reast cancer. *PloS one* 10 (7), e0130760. doi:10.1371/journal.pone.0130760

Gregory, A. D., Kliment, C. R., Metz, H. E., Kim, K. H., Kargl, J., Agostini, B. A., et al. (2015). Neutrophil elastase promotes myofibroblast differentiation in lung fibrosis. *J. Leukoc. Biol.* 98 (2), 143–152. doi:10.1189/jlb.3HI1014-493R

Gryfe, R., Kim, H., Hsieh, E. T., Aronson, M. D., Holowaty, E. J., Bull, S. B., et al. (2000). Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer. *N. Engl. J. Med.* 342 (2), 69–77. doi:10.1056/NEJM200001133420201

Gschwandtner, M., Strutzmann, E., Teixeira, M. M., Anders, H. J., Diedrichs-Möhring, M., Gerlza, T., et al. (2017). Glycosaminoglycans are important mediators of neutrophilic inflammation *in vivo*. *Cytokine* 91, 65–73. doi:10.1016/j.cyto.2016.12.008

Guay, S. P., Brisson, D., Mathieu, P., Bossé, Y., Gaudet, D., and Bouchard, L. (2015). A study in familial hypercholesterolemia suggests reduced methylomic plasticity in men with coronary artery disease. *Epigenomics* 7 (1), 17–34. doi:10.2217/epi.14.64

Hamai, K., Iwamoto, H., Ishikawa, N., Horimasu, Y., Masuda, T., Miyamoto, S., et al. (2016). Comparative study of circulating MMP-7, CCL18, KL-6, SP-A, and SP-D as disease markers of idiopathic pulmonary fibrosis. *Dis. Markers* 2016, 4759040. doi:10.1155/2016/4759040

Heukels, P., van Hulst, J. A. C., van Nimwegen, M., Boorsma, C. E., Melgert, B. N., von der Thusen, J. H., et al. (2019). Enhanced Bruton's tyrosine kinase in B-cells and autoreactive IgA in patients with idiopathic pulmonary fibrosis. *Respir. Res.* 20 (1), 232. doi:10.1186/s12931-019-1195-7

Hewlett, J. C., Kropski, J. A., and Blackwell, T. S. (2018). Idiopathic pulmonary fibrosis: Epithelial-mesenchymal interactions and emerging therapeutic targets. *Matrix Biol.* 71-72, 112–127. doi:10.1016/j.matbio.2018.03.021

Horton, W. B., and Barrett, E. J. (2021). Microvascular dysfunction in diabetes mellitus and cardiometabolic disease. *Endocr. Rev.* 42 (1), 29–55. doi:10.1210/endrev/bnaa025

Hou, J., Shi, J., Chen, L., Lv, Z., Chen, X., Cao, H., et al. (2018). M2 macrophages promote myofibroblast differentiation of LR-MSCs and are associated with pulmonary fibrogenesis. *Cell Commun. Signal* 16 (1), 89. doi:10.1186/s12964-018-0300-8

Huang, W. L., Wu, S. F., Huang, X., and Zhou, S. (2022). Integrated analysis of ECT2 and COL17A1 as potential biomarkers for pancreatic cancer. *Dis. Markers* 2022, 9453549. doi:10.1155/2022/9453549

Huang, X., Xiu, H., Zhang, S., and Zhang, G. (2018). The role of macrophages in the pathogenesis of ALI/ARDS. *Mediat. Inflamm.* 2018, 1264913. doi:10.1155/2018/1264913

Jara, P., Calyeca, J., Romero, Y., Plácido, L., Yu, G., Kaminski, N., et al. (2015). Matrix metalloproteinase (MMP)-19-deficient fibroblasts display a profibrotic phenotype. *Am. J. Physiol. Lung Cell Mol. Physiol.* 308 (6), L511–L522. doi:10.1152/ajplung.00043.2014

Jayadev, R., Morais, M., Ellingford, J. M., Srinivasan, S., Naylor, R. W., Lawless, C., et al. (2022). A basement membrane discovery pipeline uncovers network complexity, regulators, and human disease associations. *Sci. Adv.* 8 (20), eabn2265. doi:10.1126/sciadv.abn2265

Jiang, Y., Chen, F., Ren, X., Yang, Y., Luo, J., Yuan, J., et al. (2022). RNA-binding protein COL14A1, TNS1, NUSAP1 and YWHAE are valid biomarkers to predict peritoneal metastasis in gastric cancer. *Front. Oncol.* 12, 830688. doi:10.3389/fonc.2022.830688

Khan, F. A., Stewart, I., Saini, G., Robinson, K. A., and Jenkins, R. G. (2022). A systematic review of blood biomarkers with individual participant data meta-analysis of matrix metalloproteinase-7 in idiopathic pulmonary fibrosis. *Eur. Respir. J.* 59 (4), 2101612. doi:10.1183/13993003.01612-2021

King, T. E., Jr., Bradford, W. Z., Castro-Bernardini, S., Fagan, E. A., Glaspole, I., Glassberg, M. K., et al. (2014). A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N. Engl. J. Med.* 370 (22), 2083–2092. doi:10.1056/NEJMoa1402582

Kiyokawa, H., and Morimoto, M. (2020). Notch signaling in the mammalian respiratory system, specifically the trachea and lungs, in development, homeostasis, regeneration, and disease. *Dev. Growth Differ.* 62 (1), 67–79. doi:10.1111/dgd.12628

Kondoh, Y., Suda, T., Hongo, Y., Yoshida, M., Hiroi, S., Iwasaki, K., et al. (2022). Prevalence of idiopathic pulmonary fibrosis in Japan based on a claims database analysis. *Respir. Res.* 23 (1), 24. doi:10.1186/s12931-022-01938-6

Kotsianidis, I., Nakou, E., Bouchliou, I., Tzouvelekis, A., Spanoudakis, E., Steiropoulos, P., et al. (2009). Global impairment of CD4+CD25+FOXP3+ regulatory T cells in idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* 179 (12), 1121–1130. doi:10.1164/rccm.200812-1936OC

Kozawa, K., Sekai, M., Ohba, K., Ito, S., Sako, H., Maruyama, T., et al. (2021). The CD44/COL17A1 pathway promotes the formation of multilayered, transformed epithelia. *Curr. Biol.* 31 (14), 3086–3097.e7. doi:10.1016/j.cub.2021.04.078

Kruegel, J., and Miosge, N. (2010). Basement membrane components are key players in specialized extracellular matrices. *Cell Mol. Life Sci.* 67 (17), 2879–2895. doi:10.1007/s00018-010-0367-x

Kyprianou, C., Christodoulou, N., Hamilton, R. S., Nahaboo, W., Boomgaard, D. S., Amadei, G., et al. (2020). Basement membrane remodelling regulates mouse embryogenesis. *Nature* 582 (7811), 253–258. doi:10.1038/s41586-020-2264-2

Lee, K. H., Shin, S. J., Kim, K. O., Kim, M. K., Hyun, M. S., Kim, T. N., et al. (2006). Relationship between E-cadherin, matrix metalloproteinase-7 gene expression and clinicopathological features in gastric carcinoma. *Oncol. Rep.* 16 (4), 823–830. doi:10.3892/or.16.4.823

Li, H., Shen, X., Ma, M., Liu, W., Yang, W., Wang, P., et al. (2021). ZIP10 drives osteosarcoma proliferation and chemoresistance through ITGA10-mediated activation of the PI3K/AKT pathway. *J. Exp. Clin. Cancer Res.* 40 (1), 340. doi:10.1186/s13046-021-02146-8

Liao, H. Y., Da, C. M., Liao, B., and Zhang, H. H. (2021). Roles of matrix metalloproteinase-7 (MMP-7) in cancer. *Clin. Biochem.* 92, 9–18. doi:10.1016/j.clinbiochem.2021.03.003

Lindgren, O., Le Menn, G., Tuusa, J., Chen, Z. J., Tasanen, K., and Kokkonen, N. (2023). Absence of NC14A domain of COLXVII/BP180 in mice results in IL-17–associated skin inflammation. *J. Invest. Dermatol* 143 (1), 48–56.e7. doi:10.1016/j.jid.2022.07.019

Lo Re, S., Lecocq, M., Uwambayinema, F., Yakoub, Y., Delos, M., Demoulin, J. B., et al. (2011). Platelet-derived growth factor-producing CD4+ Foxp3+ regulatory T lymphocytes promote lung fibrosis. *Am. J. Respir. Crit. Care Med.* 184 (11), 1270–1281. doi:10.1164/rccm.201103-0516OC

Lothong, M., Sakares, W., Rojsitthisak, P., Tanikawa, C., Matsuda, K., and Yodsurang, V. (2021). Collagen XVII inhibits breast cancer cell proliferation and growth through deactivation of the AKT/mTOR signaling pathway. *PLoS One* 16 (7), e0255179. doi:10.1371/journal.pone.0255179

Luo, Y., Yi, H., Huang, X., Lin, G., Kuang, Y., Guo, Y., et al. (2021). Inhibition of macrophage migration inhibitory factor (MIF) as a therapeutic target in bleomycin-induced pulmonary fibrosis rats. *Am. J. Physiol. Lung Cell Mol. Physiol.* 321 (1), L6–l16. doi:10.1152/ajplung.00288.2020

Mahalanobish, S., Saha, S., Dutta, S., and Sil, P. C. (2020). Matrix metalloproteinase: An upcoming therapeutic approach for idiopathic pulmonary fibrosis. *Pharmacol. Res.* 152, 104591. doi:10.1016/j.phrs.2019.104591

Maher, T. M., Bendstrup, E., Dron, L., Langley, J., Smith, G., Khalid, J. M., et al. (2021). Global incidence and prevalence of idiopathic pulmonary fibrosis. *Respir. Res.* 22 (1), 197. doi:10.1186/s12931-021-01791-z

Mak, K. M., and Mei, R. (2017). Basement membrane type IV collagen and laminin: An overview of their biology and value as fibrosis biomarkers of liver disease. *Anat. Rec. Hob.* 300 (8), 1371–1390. doi:10.1002/ar.23567

Malta, T. M., Sokolov, A., Gentles, A. J., Burzykowski, T., Poisson, L., Weinstein, J. N., et al. (2018). Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* 173 (2), 338–354.e15. doi:10.1016/j.cell.2018.03.034

Matsuhira, T., Nishiyama, O., Tabata, Y., Kaji, C., Kubota-Ishida, N., Chiba, Y., et al. (2020). A novel phosphodiesterase 4 inhibitor, AA6216, reduces macrophage activity and fibrosis in the lung. *Eur. J. Pharmacol.* 885, 173508. doi:10.1016/j.ejphar.2020.173508

Mertens-Walker, I., Fernandini, B. C., Maharaj, M. S. N., Rockstroh, A., Nelson, C. C., Herington, A. C., et al. (2015). The tumour-promoting receptor tyrosine kinase, EphB4, regulates expres sion of integrin-β8 in prostate cancer cells. *BMC cancer* 15, 164. doi:10.1186/s12885-015-1164-6

Morris, M. R., Ricketts, C., Gentle, D., Abdulrahman, M., Clarke, N., Brown, M., et al. (2010). Identification of candidate tumour suppressor genes frequently methylated in renal cell carcinoma. *Oncogene* 29 (14), 2104–2117. doi:10.1038/onc.2009.493

Nakagawa, H., Takano, K., and Kuzumaki, H. (1999). A 16-kDa fragment of collagen type XIV is a novel neutrophil chemotactic factor purified from rat granulation tissue. *Biochem. Biophys. Res. Commun.* 256 (3), 642–645. doi:10.1006/bbrc.1999.0393

Natsuga, K., Watanabe, M., Nishie, W., and Shimizu, H. (2019). Life before and beyond blistering: The role of collagen XVII in epidermal physiology. *Exp. Dermatol* 28 (10), 1135–1141. doi:10.1111/exd.13550

Nishie, W. (2020). Collagen XVII processing and blistering skin diseases. *Acta Derm. Venereol.* 100 (5), adv00054. doi:10.2340/00015555-3399

Niu, J., Li, X. M., Wang, X., Liang, C., Zhang, Y. D., Li, H. Y., et al. (2019). DKK1 inhibits breast cancer cell migration and invasion through suppression of β-catenin/MMP7 signaling pathway. *Cancer Cell Int.* 19, 168. doi:10.1186/s12935-019-0883-1

Okada, T., Lee, A. Y., Qin, L. X., Agaram, N., Mimae, T., Shen, Y., et al. (2016). Integrin-α10 dependency identifies RAC and RICTOR as therapeutic targets in high-grade myxofibrosarcoma. *Cancer Discov.* 6 (10), 1148–1165. doi:10.1158/2159-8290.CD-15-1481

Peljto, A. L., Zhang, Y., Fingerlin, T. E., Ma, S. F., Garcia, J. G., Richards, T. J., et al. (2013). Association between the MUC5B promoter polymorphism and survival in patients with idiopathic pulmonary fibrosis. *Jama* 309 (21), 2232–2239. doi:10.1001/jama.2013.5827

Peng, X., Moore, M. W., Peng, H., Sun, H., Gan, Y., Homer, R. J., et al. (2014). CD4+CD25+FoxP3+ Regulatory Tregs inhibit fibrocyte recruitment and fibrosis via suppression of FGF-9 production in the TGF-β1 exposed murine lung. *Front. Pharmacol.* 5, 80. doi:10.3389/fphar.2014.00080

Raghu, G., Richeldi, L., Jagerschmidt, A., Martin, V., Subramaniam, A., Ozoux, M. L., et al. (2018). Idiopathic pulmonary fibrosis: Prospective, case-controlled study of natural history and circulating biomarkers. *Chest* 154 (6), 1359–1370. doi:10.1016/j.chest.2018.08.1083

Richeldi, L., Collard, H. R., and Jones, M. G. (2017). Idiopathic pulmonary fibrosis. *Lancet* 389 (10082), 1941–1952. doi:10.1016/S0140-6736(17)30866-8

Richeldi, L., du Bois, R. M., Raghu, G., Azuma, A., Brown, K. K., Costabel, U., et al. (2014). Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N. Engl. J. Med.* 370 (22), 2071–2082. doi:10.1056/NEJMoa1402584

Rousselle, P., Montmasson, M., and Garnier, C. (2019). Extracellular matrix contribution to skin wound re-epithelialization. *Matrix Biol.* 75-76, 12–26. doi:10.1016/j.matbio.2018.01.002

Saftencu, M., Braicu, C., Cojocneanu, R., Buse, M., Irimie, A., Piciu, D., et al. (2019). Gene expression patterns unveil new insights in papillary thyroid cancer. *Med. Kaunas.* 55 (8), 500. doi:10.3390/medicina55080500

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 (2), 202–206. doi:10.1038/s41588-018-0312-8

Sanli, M., Akar, E., Pehlivan, S., Bakır, K., Tuncozgur, B., Isik, A. F., et al. (2013). The relationship of metalloproteinase gene polymorphisms and lung cancer. *J. Surg. Res.* 183 (2), 517–523. doi:10.1016/j.jss.2013.01.045

Schuppan, D., Cantaluppi, M. C., Becker, J., Veit, A., Bunte, T., Troyer, D., et al. (1990). Undulin, an extracellular matrix glycoprotein associated with collagen fibrils. *J. Biol. Chem.* 265 (15), 8823–8832. doi:10.1016/s0021-9258(19)38962-8

Shapouri-Moghaddam, A., Mohammadian, S., Vazini, H., Taghadosi, M., Esmaeili, S. A., Mardani, F., et al. (2018). Macrophage plasticity, polarization, and function in health and disease. *J. Cell Physiol.* 233 (9), 6425–6440. doi:10.1002/jcp.26429

Song, J. W., Do, K. H., Jang, S. J., Colby, T. V., Han, S., and Kim, D. S. (2013). Blood biomarkers MMP-7 and SP-A: Predictors of outcome in idiopathic pulmonary fibrosis. *Chest* 143 (5), 1422–1429. doi:10.1378/chest.11-2735

Strieter, R. M., and Mehrad, B. (2009). New mechanisms of pulmonary fibrosis. *Chest* 136 (5), 1364–1370. doi:10.1378/chest.09-0510

Su, Y.-J., Lin, W.-H., Chang, Y.-W., Wei, K.-C., Liang, C.-L., Chen, S.-C., et al. (2015). Polarized cell migration induces cancer type-specific CD133/integrin/S rc/Akt/GSK3β/β-catenin signaling required for maintenance of cancer st em cell properties. *Oncotarget* 6 (35), 38029–38045. doi:10.18632/oncotarget.5703

Sun, M., Chen, Y., Liu, X., and Cui, Y. (2021). LncRNACASC9 promotes proliferation, metastasis, and cell cycle inovarian carcinoma cells through cyclinG1/TP53/MMP7 signaling. *Bioengineered* 12 (1), 8006–8019. doi:10.1080/21655979.2021.1981795

Tabuena, R. P., Nagai, S., Tsutsumi, T., Handa, T., Minoru, T., Mikuniya, T., et al. (2005). Cell profiles of bronchoalveolar lavage fluid as prognosticators of idiopathic pulmonary fibrosis/usual interstitial pneumonia among Japanese Patients. *Respiration* 72 (5), 490–498. doi:10.1159/000087673

Tamborero, D., Rubio-Perez, C., Muiños, F., Sabarinathan, R., Piulats, J. M., Muntasell, A., et al. (2018). A pan-cancer landscape of interactions between solid tumors and infiltrating immune cell populations. *Clin. Cancer Res.* 24 (15), 3717–3728. doi:10.1158/1078-0432.CCR-17-3509

Thangavelu, P. U., Krenács, T., Dray, E., and Duijf, P. H. (2016). In epithelial cancers, aberrant COL17A1 promoter methylation predicts its misexpression and increased invasion. *Clin. Epigenetics* 8, 120. doi:10.1186/s13148-016-0290-6

Tzouvelekis, A., Gomatou, G., Bouros, E., Trigidou, R., Tzilas, V., and Bouros, D. (2019). Common pathogenic mechanisms between idiopathic pulmonary fibrosis and lung cancer. *Chest* 156 (2), 383–391. doi:10.1016/j.chest.2019.04.114

van der Velden, J. L., Ye, Y., Nolin, J. D., Hoffman, S. M., Chapman, D. G., Lahue, K. G., et al. (2016). JNK inhibition reduces lung remodeling and pulmonary fibrotic systemic markers. *Clin. Transl. Med.* 5 (1), 36. doi:10.1186/s40169-016-0117-2

Van Doren, S. R. (2022). MMP-7 marks severe pancreatic cancer and alters tumor cell signaling by proteolytic release of ectodomains. *Biochem. Soc. Trans.* 50 (2), 839–851. doi:10.1042/BST20210640

Vracko, R. (1974). Basal lamina scaffold-anatomy and significance for maintenance of orderly tissue structure. *Am. J. Pathol.* 77 (2), 314–346.

Warheit-Niemi, H. I., Edwards, S. J., SenGupta, S., Parent, C. A., Zhou, X., O'Dwyer, D. N., et al. (2022). Fibrotic lung disease inhibits immune responses to staphylococcal pneumonia via impaired neutrophil and macrophage function. *JCI Insight* 7 (4), e152690. doi:10.1172/jci.insight.152690

Weis-Müller, B. T., Modlich, O., Drobinskaya, I., Unay, D., Huber, R., Bojar, H., et al. (2006). Gene expression in acute stanford type A dissection: A comparative microarray study. *J. Transl. Med.* 4, 29. doi:10.1186/1479-5876-4-29

West, J. B., and Mathieu-Costello, O. (1999). Structure, strength, failure, and remodeling of the pulmonary blood-gas barrier. *Annu. Rev. Physiol.* 61, 543–572. doi:10.1146/annurev.physiol.61.1.543

Wilson, S. E. (2020). Corneal wound healing. *Exp. Eye Res.* 197, 108089. doi:10.1016/j.exer.2020.108089

Wilson, S. E. (2022). Fibrosis is a basement membrane-related disease in the cornea: Injury and defective regeneration of basement membranes may underlie fibrosis in other organs. *Cells* 11 (2), 309. doi:10.3390/cells11020309

Wolters, P. J., Collard, H. R., and Jones, K. D. (2014). Pathogenesis of idiopathic pulmonary fibrosis. *Annu. Rev. Pathol.* 9, 157–179. doi:10.1146/annurev-pathol-012513-104706

Wynn, T. A., and Vannella, K. M. (2016). Macrophages in tissue repair, regeneration, and fibrosis. *Immunity* 44 (3), 450–462. doi:10.1016/j.immuni.2016.02.015

Xiong, S., Guo, R., Yang, Z., Xu, L., Du, L., Li, R., et al. (2015). Treg depletion attenuates irradiation-induced pulmonary fibrosis by reducing fibrocyte accumulation, inducing Th17 response, and shifting IFN-γ, IL-12/IL-4, IL-5 balance. *Immunobiology* 220 (11), 1284–1291. doi:10.1016/j.imbio.2015.07.001

Xu, F., Tanabe, N., Vasilescu, D. M., McDonough, J. E., Coxson, H. O., Ikezoe, K., et al. (2021). The transition from normal lung anatomy to minimal and established fibrosis in idiopathic pulmonary fibrosis (IPF). *EBioMedicine* 66, 103325. doi:10.1016/j.ebiom.2021.103325

Yang, D., Xu, P., Su, H., Zhong, W., Xu, J., Su, Z., et al. (2022). The histone methyltransferase DOT1L is a new epigenetic regulator of pulmonary fibrosis. *Cell Death Dis.* 13 (1), 60. doi:10.1038/s41419-021-04365-5

Yodsurang, V., Tanikawa, C., Miyamoto, T., Lo, P. H. Y., Hirata, M., and Matsuda, K. (2017). Identification of a novel p53 target, COL17A1, that inhibits breast cancer cell migration and invasion. *Oncotarget* 8 (34), 55790–55803. doi:10.18632/oncotarget.18433

Yuan, S., Lin, L. S., Gan, R. H., Huang, L., Wu, X. T., Zhao, Y., et al. (2020). Elevated matrix metalloproteinase 7 expression promotes the proliferation, motility and metastasis of tongue squamous cell carcinoma. *BMC Cancer* 20 (1), 33. doi:10.1186/s12885-020-6521-4

Zolak, J. S., and de Andrade, J. A. (2012). Idiopathic pulmonary fibrosis. *Immunol. Allergy Clin. North Am.* 32 (4), 473–485. doi:10.1016/j.iac.2012.08.006

Zuo, F., Kaminski, N., Eugui, E., Allard, J., Yakhini, Z., Ben-Dor, A., et al. (2002). Gene expression analysis reveals matrilysin as a key regulator of pulmonary fibrosis in mice and humans. *Proc. Natl. Acad. Sci. U. S. A.* 99 (9), 6292–6297. doi:10.1073/pnas.092134099

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 20 March 2023
DOI 10.3389/fgene.2023.1164274

Check for updates

OPEN ACCESS

EDITED BY
Xianyong Yin,
Department of Biostatistics, University of Michigan, Ann Arbor, United States

REVIEWED BY
Xingjie Shi,
East China Normal University, China
Yue Fan,
Xi'an Jiaotong University, China

*CORRESPONDENCE
Rongbin Yu,
rongbinyu@njmu.edu.cn
Peng Huang,
huangpeng@njmu.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

RECEIVED 12 February 2023
ACCEPTED 03 March 2023
PUBLISHED 20 March 2023

CITATION
Zhu X, Zou Y, Jia L, Ye X, Zou Y, Tu J, Li J, Yu R, Yang S and Huang P (2023), Using multi-tissue transcriptome-wide association study to identify candidate susceptibility genes for respiratory infectious diseases.
Front. Genet. 14:1164274.
doi: 10.3389/fgene.2023.1164274

COPYRIGHT
© 2023 Zhu, Zou, Jia, Ye, Zou, Tu, Li, Yu, Yang and Huang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Using multi-tissue transcriptome-wide association study to identify candidate susceptibility genes for respiratory infectious diseases

Xiaobo Zhu[1†], Yixin Zou[2†], Linna Jia[2†], Xiangyu Ye[2], Yanzheng Zou[2], Junlan Tu[2], Juntong Li[2], Rongbin Yu[2*], Sheng Yang[3] and Peng Huang[2*]

[1]The People's Hospital of Danyang, Affiliated Danyang Hospital of Nantong University, Zhenjiang, China, [2]Department of Epidemiology, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, China, [3]Department of Biostatistics, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, China

Objective: We explore the candidate susceptibility genes for influenza A virus (IAV), measles, rubella, and mumps and their underlying biological mechanisms.

Methods: We downloaded the genome-wide association study summary data of four virus-specific immunoglobulin G (IgG) level data sets (anti-IAV IgG, anti-measles IgG, anti-rubella IgG, and anti-mumps virus IgG levels) and integrated them with reference models of three potential tissues from the Genotype-Tissue Expression (GTEx) project, namely, whole blood, lung, and transformed fibroblast cells, to identify genes whose expression is predicted to be associated with IAV, measles, mumps, and rubella.

Results: We identified 19 significant genes (ULK4, AC010132.11, SURF1, NIPAL2, TRAP1, TAF1C, AC000078.5, RP4-639F20.1, RMDN2, ATP1B3, SRSF12, RP11-477D19.2, TFB1M, XXyac-YX65C7_A.2, TAF1C, PCGF2, and BNIP1) associated with IAV at a Bonferroni-corrected threshold of $p < 0.05$; 14 significant genes (SOAT1, COLGALT2, AC021860.1, HCG11, METTL21B, MRPL10, GSTM4, PAQR6, RP11-617D20.1, SNX8, METTL21B, ANKRD27, CBWD2, and TSFM) associated with measles at a Bonferroni-corrected threshold of $p < 0.05$; 15 significant genes (MTOR, LAMC1, TRIM38, U91328.21, POLR2J, SCRN2, Smpd4, UBN1, CNTROB, SCRN2, HOXB-AS1, SLC14A1, AC007566.10, AC093668.2, and CPD) associated with mumps at a Bonferroni-corrected threshold of $p < 0.05$; and 13 significant genes (JAGN1, RRP12, RP11-452K12.7, CASP7, AP3S2, IL17RC, FAM86HP, AMACR, RRP12, PPP2R1B, C11orf1, DLAT, and TMEM117) associated with rubella at a Bonferroni-corrected threshold of $p < 0.05$.

Conclusions: We have identified several candidate genes for IAV, measles, mumps, and rubella in multiple tissues. Our research may further our understanding of the pathogenesis of infectious respiratory diseases.

KEYWORDS

transcriptome-wide association study, IAV, measles, mumps, rubella, multiple tissues, genome-wide association studies, gene expression

Introduction

Although respiratory infections are largely preventable causes of illness and death, they remain the leading cause of death from infectious diseases worldwide, ranking fifth among all total causes of death (GBD 2015 LRI Collaborators, 2017). Influenza A virus (IAV), measles virus, rubella virus, and mumps virus infect a wide range of hosts, are highly transmissible and are prone to a mutation that can cause pandemics in a short period of time, placing an enormous burden and pressure on public health systems (Marshall and Plotkin, 2019; Harrington et al., 2021; Iacobucci, 2022; Raghunathan and Orenstein, 2022). The incidence of IAV, measles, mumps, and rubella has decreased drastically following the implementation of vaccination programs. However, the segmented genome of IAV allows for easy recombination and antigenic shift, resulting in novel antigens (Bouvier and Palese, 2008). Thus, the ability of IAV to rapidly evolve can result in highly pathogenic viral strains. Measles and rubella remain endemic in many countries, leading to the importation of cases and occasional local transmission within China, and the resurgence of mumps has recently been reported, with outbreaks still occurring and challenges remaining while controlling these diseases (Kauffmann et al., 2021). There is considerable variation in the severity of respiratory infections caused by viral infections. There are major determinants of this variability, such as intrinsic virus pathogenicity, acquired host factors (e.g., immunity and coexisting conditions), and innate host susceptibility. While viral genetic determinants of respiratory disease severity and host immunity have been well studied, host genetic determinants have been much less explored (Kennedy et al., 2014; Voigt et al., 2018; Sabikunnahar et al., 2022).

Genome-wide association studies (GWAS) are effective methods for understanding the genetic basis of many complex traits in common human diseases (Hayes, 2013). In particular, these have proven to be well-suited for identifying common single-nucleotide polymorphism (SNP) variants with moderate to large effects on phenotypes (Gorlova et al., 2022). However, the specific biological mechanisms and functional consequences of many of the genetic variants identified by the GWAS remain unclear, especially their role in disease severity; that is, GWAS methods may miss small effect–trait associations. Gene expression is an intermediate phenotype between genetic variation and an underlying disease predisposition trait (Albert and Kruglyak, 2015). Many genetic variants affect complex traits by regulating gene expressions. Unfortunately, large-scale expression–trait associations are hampered by sample availability and cost, as well as intrinsic factors and small effects. Therefore, to address these issues, transcriptome-wide association studies (TWASs) were developed, which integrate gene expression into large-scale GWAS (Derks and Gamazon, 2020). Through extensive simulation of existing GWAS data, TWASs have identified candidate genes associated with mental disorders (Thériault et al., 2018), calcified aortic stenosis (Liu et al., 2020), pancreatic cancer (Lin et al., 2017), and inflammatory biological age (Zhang et al., 2022). However, for many complex traits, biologically relevant tissues are unknown. Most existing studies identify gene–trait associations on the basis of a single tissue, whereby the significant gene effects that are identified get inflated. In addition, it has been shown that eQTLs with larger effects tend to regulate gene expressions in multiple tissues. The TWAS analysis of multiple tissues improves the accuracy of the results.

There have been some studies exploring the association of respiratory infectious diseases with some human tissues. Pulmonary inflammation and airway epithelial injury are hallmarks of human pulmonary infectious diseases. The association of HLA alleles with respiratory infectious diseases in lung cells has been confirmed by Zhang et al. (2022). Moreover, Xiao et al. (2023) found that interleukin-11 (IL-11) as a fibrotic factor may be associated with respiratory diseases leading to respiratory failure or even death. These potential connections deserve our attention.

In this study, we used TWASs to explore genes associated with IAV, measles, rubella, and mumps. Given that respiratory infectious diseases are associated with multiple tissues, we used eQTL reference panels for three tissues, namely, whole blood, lung, and transformed fibroblast cells, and pooled the GWAS data for IAV, measles, rubella, and mumps to identify tissue-specific susceptibility genes. This study lays the foundation for further understanding the pathogenesis of these four viruses.

Methods

GWAS data collection and processing

We collected four virus-specific immunoglobulin G (IgG) level data sets (anti-IAV IgG, anti-measles IgG, anti-rubella IgG, and anti-mumps virus IgG levels) from the NHGRI-EBI GWAS directory (https://www.ebi.ac.uk/gwas/downloads/summary-statistics). These data sets were originally obtained from 1,000 healthy individuals (MI), and their serum prevalence rates were 77.7%, 88.5%, 93.5%, and 91.2%, respectively, for these four viruses (Scepanovic et al., 2018). Considering that the main unit of anti-IAV IgG quantitative unit is S/CO (signal/cut-off ratio), the main unit of anti-measles and mumps virus IgG level is IA (index antibody), and the main unit of anti-rubella virus IgG is UI/mL. We also use SD as the change unit. Then, we screened the above data for single-nucleotide polymorphisms (SNPs) for further analysis. We retained the following SNPs: 1) autosomes (1–22), 2) minimum allele frequencies (MAF) greater than 0.001, and 3) more than 70% of observers having these specific SNPs (Yang and Zhou, 2020; Yang and Zhou, 2022). We also plotted Manhattan plots and q-q plots for the processed GWAS data sets for four traits to show the frequency and distribution of their genes (Figure 1; Supplementary Figure S1).

Transcriptome-wide association analysis

We used FUSION to perform relevant TWAS tissue analysis using GWAS summary data formatted after the processing of whole blood, lung, and transformed fibroblast cells (Gusev et al., 2016). Specifically, FUSION uses the pre-calculated gene expression weight to perform summary statistics with respiratory infectious diseases GWAS and calculates the association between each gene and respiratory infectious diseases. Association statistics are defined as TWAS z-scores and estimated as follows:

$$Z_{TWAS} = WZ \Big/ \left(W \sum_{S,S} W^{t}\right)^{1/2}.$$

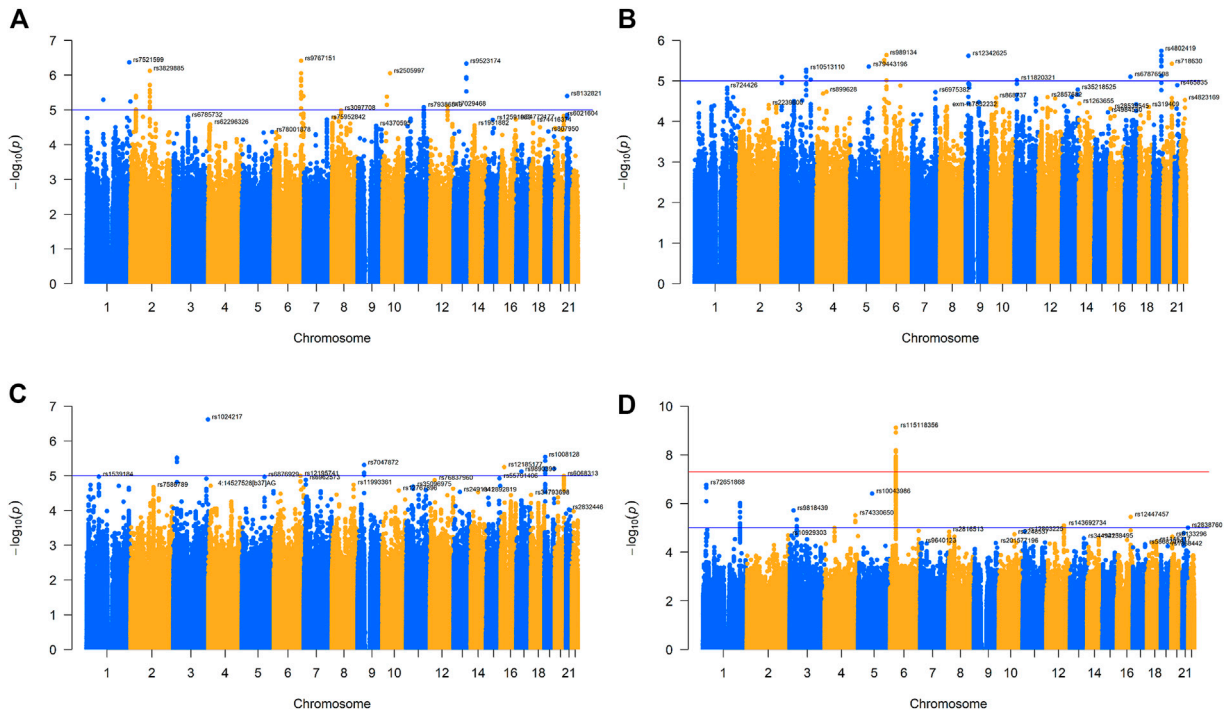


FIGURE 1
Manhattan plot for four GWAS of respiratory infectious diseases. **(A)**. Manhattan plot of IAV. **(B)**. Manhattan plot of measles. **(C)**. Manhattan plot of mumps. **(D)**. Manhattan plot of rubella.

The weight of Z and W plays a crucial role in the distribution of Z_{TWAS}. In the formula, Z is the estimated value of the SNPs for the disease under study, w^t is the weight, and s is the linkage disequilibrium between all SNPs. It is worth noting that Z_{TWAS} is only well calibrated under the zero model of $Z \sim N(0, \Sigma s,s)$ (mean 0, unit variance 0). The reference panel that we used came from the 1000 Genomes Project in Europe (Speed et al., 2020). We also downloaded from FUSION (http://gusevlab.org/projects/fusion/) pre-computed gene expression weights based on GTEx v7 feature weights and used these for the research of each tissue. Based on the highest cross-validation R^2, FUSION selects the best expression prediction model among the top eQTL, the best linear predictor (BLUP) (Gusev et al., 2016), the Bayesian linear mixed model (BSLMM) (Zhou et al., 2013), the least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996; Veturi and Ritchie, 2018), and the elastic net (Hui and Hastie, 2005) to calculate the SNP expression weight within a 1 Mb window for a given gene. We also applied the Bonferroni correction to explain multiple hypotheses and set the p-value threshold to 0.01.

Results

Multi-tissue transcriptome-wide significant genes of IAV

We used FUSION to assess the relationship between the predictive gene expression and IAV. After Bonferroni correction, we found that 19 genes in the 3 tissues were significantly associated with IAV (Figure 2A; Table 1). Specifically, in the whole blood, we found ULK4 ($p_{adjusted}$ = 0.012), AC010132.11 ($p_{adjusted}$ = 0.022), and SURF1 ($p_{adjusted}$ = 0.030) to be IAV-related genes, while in the lung tissue, we found NIPAL2 ($p_{adjusted}$ = 0.019), TRAP1 ($p_{adjusted}$ = 0.042), TAF1C ($p_{adjusted}$ = 0.018), and AC000078.5 ($p_{adjusted}$ = 0.020) to be significantly associated with IAV. Other significant genes included RP4-639F20.1 ($p_{adjusted}$ = 0.041), RMDN2 ($p_{adjusted}$ = 0.009), ATP1B3 ($p_{adjusted}$ = 0.050), SRSF12 ($p_{adjusted}$ = 0.033), RP11-477D19.2 ($p_{adjusted}$ = 0.007), TFB1M ($p_{adjusted}$ = 0.015), XXyac-YX65C7_A.2 ($p_{adjusted}$ = 0.021), TAF1C ($p_{adjusted}$ = 0.030), and PCGF2 ($p_{adjusted}$ = 0.027) in the transformed fibroblast cells. However, we found that BNIP1 expressed in the whole blood, lung tissue, and transformed fibroblast cells was significantly associated with IAV.

Multi-tissue transcriptome-wide significant genes of measles

We used FUSION to assess the relationship between predictive gene expression and measles. After Bonferroni correction, we found that 14 genes in the three tissues were significantly associated with measles (Figure 2B; Table 2). Specifically, in the whole blood, we found SOAT1 ($p_{adjusted}$ = 0.018), COLGALT2 ($p_{adjusted}$ = 0.000), AC021860.1 ($p_{adjusted}$ = 0.044), HCG11 ($p_{adjusted}$ = 0.000), METTL21B ($p_{adjusted}$ = 0.014), and MRPL10 ($p_{adjusted}$ = 0.049) as measles-related genes, while in the lung tissue, we found GSTM4 ($p_{adjusted}$ = 0.017), PAQR6 ($p_{adjusted}$ = 0.031), RP11-617D20.1 ($p_{adjusted}$ = 0.024), SNX8 ($p_{adjusted}$ = 0.009), METTL21B ($p_{adjusted}$ =

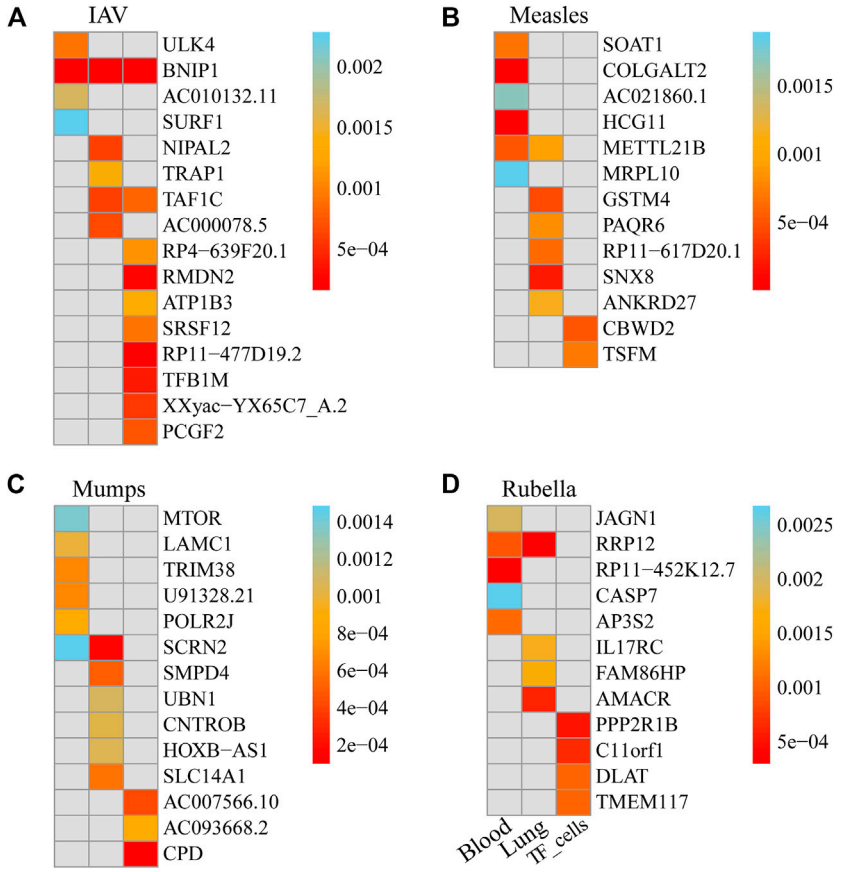


FIGURE 2
Genes significantly associated with the risk of four respiratory infectious diseases in each tissue. **(A)**. Genes significantly associated with the risk of IAV in each tissue. **(B)**. Genes significantly associated with the risk of measles in each tissue. **(C)**. Genes significantly associated with the risk of mumps in each tissue. **(D)**. Genes significantly associated with the risk of rubella in each tissue.

0.035), and ANKRD27 ($p_{adjusted}$ = 0.044) to be significantly associated with measles. Other significant genes included CBWD2 ($p_{adjusted}$ = 0.029) and TSFM ($p_{adjusted}$ = 0.041) in transformed fibroblast cells.

Multi-tissue transcriptome-wide significant genes of mumps

We used FUSION to assess the relationship between predictive gene expression and mumps. After Bonferroni correction, we found that 15 genes in the three tissues were significantly associated with mumps (Figure 2C; Table 3). In mumps, we found six significant genes, namely, MTOR ($p_{adjusted}$ = 0.029), LAMC1 ($p_{adjusted}$ = 0.021), TRIM38 ($p_{adjusted}$ = 0.014), U91328.21 ($p_{adjusted}$ = 0.014), POLR2J ($p_{adjusted}$ = 0.018), and SCRN2 ($p_{adjusted}$ = 0.031) in the whole blood; similarly, in the lung tissue, we found Smpd4 ($p_{adjusted}$ = 0.022), UBN1 ($p_{adjusted}$ = 0.049), CNTROB ($p_{adjusted}$ = 0.046), SCRN2 ($p_{adjusted}$ = 0.008), HOXB-AS1 ($p_{adjusted}$ = 0.048), and SLC14A1 ($p_{adjusted}$ = 0.027) as mumps-related genes, while in the transformed fibroblast cells, we found AC007566.10 ($p_{adjusted}$ = 0.023), AC093668.2 ($p_{adjusted}$ = 0.047), and CPD ($p_{adjusted}$ = 0.005) as significantly associated with mumps.

Multi-tissue transcriptome-wide significant genes of rubella

We used FUSION to assess the relationship between predictive gene expression and rubella. After Bonferroni correction, we found that 13 genes in the three tissues were significantly associated with rubella (Figure 2D; Table 4). In rubella, we found five significant genes, namely, JAGN1 ($p_{adjusted}$ = 0.036), RRP12 ($p_{adjusted}$ = 0.017), RP11-452K12.7 ($p_{adjusted}$ = 0.006), CASP7 ($p_{adjusted}$ = 0.048), and AP3S2 ($p_{adjusted}$ = 0.020) in the whole blood, while in the lung tissue, we found IL17RC ($p_{adjusted}$ = 0.035), FAM86HP ($p_{adjusted}$ = 0.034), AMACR ($p_{adjusted}$ = 0.012), and RRP12 ($p_{adjusted}$ = 0.006) to be significantly associated with rubella. Similarly, in transformed fibroblast cells, we found PPP2R1B ($p_{adjusted}$ = 0.022), C11orf1 ($p_{adjusted}$ = 0.029), DLAT ($p_{adjusted}$ = 0.045), and TMEM117 ($p_{adjusted}$ = 0.046) as rubella-related genes.

Discussion

In this study, we used tissue-specific TWAS to explore the association of genetically predicted gene expression with IAV,

TABLE 1 Genes significantly associated with the risk of IAV.

Tissue	Gene	CHR	MODEL	TWAS.Z	TWAS.P	*P-Bonferroni*
Blood	ULK4	3	ENet	−3.319	9.050E-04	0.012
Blood	BNIP1	5	Top1	−3.763	1.680E-04	0.002
Blood	AC010132.11	7	Top1	3.142	1.680E-03	0.022
Blood	SURF1	9	LASSO	3.050	2.290E-03	0.030
Lung	BNIP1	5	LASSO	−3.730	1.910E-04	0.006
Lung	NIPAL2	8	LASSO	3.422	6.220E-04	0.019
Lung	TRAP1	16	ENet	−3.192	1.411E-03	0.042
Lung	TAF1C	16	ENet	3.431	6.010E-04	0.018
Lung	AC000078.5	22	ENet	−3.409	6.520E-04	0.020
T.F cells	RP4-639F20.1	1	LASSO	−3.252	1.140E-03	0.041
T.F cells	RMDN2	2	ENet	3.658	2.540E-04	0.009
T.F cells	ATP1B3	3	Top1	3.199	1.380E-03	0.050
T.F cells	BNIP1	5	ENet	−3.714	2.040E-04	0.007
T.F cells	SRSF12	6	Top1	3.314	9.200E-04	0.033
T.F cells	RP11-477D19.2	6	Top1	−3.731	1.910E-04	0.007
T.F cells	TFB1M	6	ENet	−3.536	4.070E-04	0.015
T.F cells	XXyac-YX65C7_A.2	6	ENet	3.443	5.760E-04	0.021
T.F cells	TAF1C	16	ENet	3.342	8.310E-04	0.030
T.F cells	PCGF2	17	LASSO	3.367	7.600E-04	0.027

CHR, the chromosome on which the identified gene is located; MODEL, models used for imputation in FUSION; TWAS.P, *p*-values for TWAS analysis in each tissue; P-Bonferroni, *p*-values corrected by Bonferroni for TWAS analysis in each tissue.

TABLE 2 Genes significantly associated with the risk of measles.

Tissue	Gene	CHR	MODEL	TWAS.Z	TWAS.P	*P-Bonferroni*
Blood	SOAT1	1	Top1	3.399	6.760E-04	0.018
Blood	COLGALT2	1	Top1	4.308	1.650E-05	0.000
Blood	AC021860.1	4	LASSO	3.137	1.710E-03	0.044
Blood	HCG11	6	LASSO	−4.459	8.220E-06	0.000
Blood	METTL21B	12	Top1	3.465	5.300E-04	0.014
Blood	MRPL10	17	ENet	−3.106	1.900E-03	0.049
Lung	GSTM4	1	ENet	−3.512	4.440E-04	0.017
Lung	PAQR6	1	ENet	3.345	8.240E-04	0.031
Lung	RP11-617D20.1	4	ENet	3.415	6.380E-04	0.024
Lung	SNX8	7	LASSO	−3.680	2.330E-04	0.009
Lung	METTL21B	12	ENet	3.314	9.210E-04	0.035
Lung	ANKRD27	19	ENet	3.252	1.150E-03	0.044
T.F cells	CBWD2	2	LASSO	3.475	5.110E-04	0.029
T.F cells	TSFM	12	ENet	−3.383	7.160E-04	0.041

CHR, the chromosome on which the identified gene is located; MODEL, models used for imputation in FUSION; TWAS.P, *p*-values for TWAS analysis in each tissue; P-Bonferroni, *p*-values corrected by Bonferroni for TWAS analysis in each tissue.

TABLE 3 Genes significantly associated with the risk of mumps.

Tissue	Gene	CHR	MODEL	TWAS.Z	TWAS.P	*P-Bonferroni*
Blood	MTOR	1	ENet	−3.195	1.400E-03	0.029
Blood	LAMC1	1	Top1	−3.289	1.010E-03	0.021
Blood	TRIM38	6	Top1	−3.401	6.710E-04	0.014
Blood	U91328.21	6	Top1	−3.401	6.710E-04	0.014
Blood	POLR2J	7	Top1	−3.327	8.780E-04	0.018
Blood	SCRN2	17	ENet	3.177	1.490E-03	0.031
Lung	Smpd4	2	LASSO	3.483	4.960E-04	0.022
Lung	UBN1	16	ENet	3.266	1.090E-03	0.049
Lung	CNTROB	17	ENet	−3.285	1.020E-03	0.046
Lung	SCRN2	17	LASSO	3.738	1.850E-04	0.008
Lung	HOXB-AS1	17	LASSO	3.274	1.060E-03	0.048
Lung	SLC14A1	18	LASSO	3.436	5.910E-04	0.027
T.F cells	AC007566.10	7	Top1	−3.524	4.250E-04	0.023
T.F cells	AC093668.2	7	Top1	3.327	8.780E-04	0.047
T.F cells	CPD	17	LASSO	−3.886	1.020E-04	0.005

CHR, the chromosome on which the identified gene is located; MODEL, models used for imputation in FUSION; TWAS.P, *p*-values for TWAS analysis in each tissue; P-Bonferroni, *p*-values corrected by Bonferroni for TWAS analysis in each tissue.

TABLE 4 Genes significantly associated with the risk of rubella.

Tissue	Gene	CHR	MODEL	TWAS.Z	TWAS.P	*P-Bonferroni*
Blood	JAGN1	3	Top1	−3.092	1.990E-03	0.036
Blood	RRP12	10	LASSO	−3.303	9.560E-04	0.017
Blood	RP11-452K12.7	10	LASSO	−3.608	3.090E-04	0.006
Blood	CASP7	10	Top1	3.002	2.682E-03	0.048
Blood	AP3S2	15	LASSO	3.263	1.100E-03	0.020
Lung	IL17RC	3	ENet	3.126	1.770E-03	0.035
Lung	FAM86HP	3	ENet	−3.136	1.710E-03	0.034
Lung	AMACR	5	ENet	3.425	6.160E-04	0.012
Lung	RRP12	10	Top1	−3.614	3.020E-04	0.006
T.F cells	PPP2R1B	11	ENet	−3.485	4.930E-04	0.022
T.F cells	C11orf1	11	ENet	3.408	6.560E-04	0.029
T.F cells	DLAT	11	Top1	−3.282	1.031E-03	0.045
T.F cells	TMEM117	12	ENet	3.277	1.050E-03	0.046

CHR, the chromosome on which the identified gene is located; MODEL, models used for imputation in FUSION; TWAS.P, *p*-values for TWAS analysis in each tissue; P-Bonferroni, *p*-values corrected by Bonferroni for TWAS analysis in each tissue.

measles, rubella, and mumps and identified 19, 14, 15, and 13 genes in three tissues as potential IAV-, measles-, mumps-, and rubella- related genes, respectively. In IAV, including three in whole blood (ULK4, AC010132.11, and SURF1), four in the lung tissue (NIPAL2, TRAP1, TAF1C, and AC000078.5), nine in the transformed fibroblast cells (RP4-639F20.1, RMDN2, ATP1B3, SRSF12, RP11-477D19.2, TFB1M, XXyac-YX65C7_A.2, TAF1C, and PCGF2), and one expressed in three tissues

as significantly associated with IAV (BNIP1). In measles, including six in whole blood (SOAT1, COLGALT2, AC021860.1, HCG11, METTL21B, and MRPL10), six in the lung tissue (GSTM4, PAQR6, RP11-617D20.1, SNX8, METTL21B, and ANKRD27), and two in the transformed fibroblast cells (CBWD2, and TSFM). In mumps, including six in whole blood (MTOR, LAMC1, TRIM38, U91328.21, POLR2J, and SCRN2), six in the lung tissue (Smpd4, UBN1, CNTROB, SCRN2, HOXB-AS1, and SLC14A1), and three in the transformed fibroblast cells (AC007566.10, AC093668.2, and CPD). In rubella, including five in whole blood (JAGN1, RRP12, RP11-452K12.7, CASP7, and AP3S2), four in the lung tissue (IL17RC, FAM86HP, AMACR, and RRP12), and four in the transformed fibroblast cells (PPP2R1B, C11orf1, DLAT, and TMEM117). The above 59 genes have never been clearly reported to be associated with IAV, measles, mumps, and rubella.

TNF receptor-associated protein-1 (TRAP1), a member of the mitochondria-specific Hsp90 family, is located in the mitochondrial matrix, mitochondrial endometrium, and the intermembrane space (Cechetto and Gupta, 2000; Pridgeon et al., 2007). However, among the few real TRAP1 clients described, two are subunits of the electron transport chain complex (ETC), a component of the complex II succinate dehydrogenase subunit A/B (SDHA/B) (Sanchez-Martin et al., 2020; Sanchez-Martin et al., 2021), and complex IV cytochromes c oxidase subunit 2 (COXII) (Xiang et al., 2016; Xiang et al., 2019). The II/SDH complex is a protein complex containing iron and sulfur groups whose function is to transfer electrons from succinate to the coenzyme Q10-ubiquinone (III complex) (Bezawork-Geleta et al., 2017). TRAP1 leaves the SDH in a partially unfolded state. Inhibition of TRAP1 can release active SDH and increase its activity (Agarwal et al., 2019; Hu et al., 2020). In addition, SDH activity (Masgras et al., 2017; Hu et al., 2020) and oxygen consumption (Masgras et al., 2017) were negatively correlated with TRAP1 expression, suggesting that TRAP1 promotes the Warburg effect (Guzzo et al., 2014). It should be noted that SDH also oxidizes succinic acid to fumarate, thus integrating the TCA cycle, suggesting a broad impact of TRAP1 on mitochondrial metabolism (Guzzo et al., 2014; Hsu et al., 2016). It can be assumed that the abnormal expression of the TRAP1 gene may contribute to the development of IAV by affecting mitochondrial metabolic processes.

GSTM4 belongs to the mu class of glutathione S-transferase (GST). Compared to other GSTMs (Comstock et al., 1993), GSTM4 exhibits greater consistency in amino acid sequence but has significant differences in physicochemical properties and tissue distribution. GSTM4 does not exhibit activity comparable to standard TPS substrates, and its specific substrate has not been identified. It is well known that the GST gene is highly polymorphic to adapt to the growing number of exotic compounds (Hayes and Strange, 2000). This polymorphism can alter individual susceptibility to the disease and response to therapeutic agents. For example, the T2517C polymorphism in GSTM4 has been shown to be associated with an increased risk of lung cancer (Liloglou et al., 2002). The mechanism underlying this association is not yet clear. Currently, there are few reports of the relationship between the GSTM4 gene and the disease. Efforts may be needed to verify its role in infectious respiratory diseases.

In addition to the aforementioned genes, we have identified other candidate genes (SOAT1, HCG11, PAQR6, and AP3S2) in different tissues, of which AP3S2 has been reported to be associated with the development of type 2 diabetes (Kazakova et al., 2017). Other genes have also been implicated in cancer. For example, SOAT1 has been linked to the development of stomach cancer (Zhu et al., 2021), and pancreatic cancer (Oni et al., 2020); HCG11 may influence the development of nasopharyngeal carcinoma (Zheng et al., 2022); the PAQR6 gene has also been associated with several cancers (Yang et al., 2021). Although there is little research on these genes and respiratory infectious diseases, the potential links between these genes deserve further attention in future studies.

In summary, our results list some genes for respiratory infections that have not been studied before. Based on the previous knowledge of these genes, this means that they may be involved in host infection through transcriptional processes of susceptible genes and RNA degradation processes. This study is helpful to deepen the understanding of the pathogenesis of respiratory infectious diseases.

Data availability statement

Publicly available data sets were analyzed in this study. These data can be found here: https://www.ebi.ac.uk/gwas/downloads/summary-statistics.

Author contributions

XZ, YZ, and LJ contributed equally to the conception, design, and execution of the study, as well as to the analysis and interpretation of data. XZ, YZ, LJ, XY, YZo, JT, JL, SY, RY, and PH drafted, revised, and approved the manuscript. RY and PH are the corresponding authors who took responsibility for the final submission. All authors agree to be accountable for all aspects of the work, ensuring the accuracy and integrity of the content.

Funding

This study was sponsored by the National Natural Science Foundation of China (82173585 and 82273741), the Natural Science Foundation of Jiangsu Province (BK20190106), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Acknowledgments

This study is a joint effort of many investigators and staff members, and their contributions are gratefully acknowledged.

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, editors, and 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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1164274/full#supplementary-material

References

Agarwal, E., Altman, B. J., Seo, J. H., Ghosh, J. C., Kossenkov, A. V., Tang, H. Y., et al. (2019). Myc-mediated transcriptional regulation of the mitochondrial chaperone TRAP1 controls primary and metastatic tumor growth. *J. Biol. Chem.* 294 (27), 10407–10414. doi:10.1074/jbc.AC119.008656

Albert, F. W., and Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* 16 (4), 197–212. doi:10.1038/nrg3891

Bezawork-Geleta, A., Rohlena, J., Dong, L., Pacak, K., and Neuzil, J. (2017). Mitochondrial complex II: At the crossroads. *Trends Biochem Sci.* 42 (4), 312–325. doi:10.1016/j.tibs.2017.01.003

Bouvier, N. M., and Palese, P. The biology of influenza viruses. *Vaccine*. 2008;26 (4): D49–D53. doi:10.1016/j.vaccine.2008.07.039

Cechetto, J. D., and Gupta, R. S. (2000). Immunoelectron microscopy provides evidence that tumor necrosis factor receptor-associated protein 1 (TRAP-1) is a mitochondrial protein which also localizes at specific extramitochondrial sites. *Exp. Cell Res.* 260 (1), 30–39. doi:10.1006/excr.2000.4983

Comstock, K. E., Johnson, K. J., Rifenbery, D., and Henner, W. D. (1993). Isolation and analysis of the gene and cDNA for a human Mu class glutathione S-transferase, GSTM4. *J. Biol. Chem.* 268 (23), 16958–16965. doi:10.1016/s0021-9258(19)85287-0

Derks, E. M., and Gamazon, E. R. (2020). Transcriptome-wide association analysis offers novel opportunities for clinical translation of genetic discoveries on mental disorders. *World Psychiatry* 19 (1), 113–114. doi:10.1002/wps.20702

GBD 2015 LRI Collaborators (2017). Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory tract infections in 195 countries: A systematic analysis for the global burden of disease study 2015. *Lancet Infect. Dis.* 17 (11), 1133–1161. doi:10.1016/S1473-3099(17)30396-1

Gorlova, O. Y., Xiao, X., Tsavachidis, S., Amos, C. I., and Gorlov, I. P. (2022). SNP characteristics and validation success in genome wide association studies. *Hum. Genet.* 141 (2), 229–238. doi:10.1007/s00439-021-02407-8

Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B. W. J. H., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* 48 (3), 245–252. doi:10.1038/ng.3506

Guzzo, G., Sciacovelli, M., Bernardi, P., and Rasola, A. (2014). Inhibition of succinate dehydrogenase by the mitochondrial chaperone TRAP1 has antioxidant and anti-apoptotic effects on tumor cells. *Oncotarget* 5 (23), 11897–11908. doi:10.18632/oncotarget.2472

Harrington, W. N., Kackos, C. M., and Webby, R. J. (2021). The evolution and future of influenza pandemic preparedness. *Exp. Mol. Med.* 53 (5), 737–749. doi:10.1038/s12276-021-00603-0

Hayes, B. (2013). Overview of statistical methods for genome-wide association studies (GWAS). *Methods Mol. Biol.* 1019, 149–169. doi:10.1007/978-1-62703-447-0_6

Hayes, J. D., and Strange, R. C. (2000). Glutathione S-transferase polymorphisms and their biological consequences. *Pharmacology* 61 (3), 154–166. doi:10.1159/000028396

Hsu, C. C., Tseng, L. M., and Lee, H. C. (2016). Role of mitochondrial dysfunction in cancer progression. *Exp. Biol. Med. (Maywood).* 241 (12), 1281–1295. doi:10.1177/1535370216641787

Hu, S., Ferraro, M., Thomas, A. P., Chung, J. M., Yoon, N. G., Seol, J. H., et al. (2020). Dual binding to orthosteric and allosteric sites enhances the anticancer activity of a TRAP1-targeting drug. *J. Med. Chem.* 63 (6), 2930–2940. doi:10.1021/acs.jmedchem.9b01420

Hui, Z., and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc.* 67 (5), 768.

Iacobucci, G. (2022). Measles is now "an imminent threat" globally, WHO and CDC warn. *BMJ* 379, o2844. doi:10.1136/bmj.o2844

Kauffmann, F., Heffernan, C., Meurice, F., Ota, M. O. C., Vetter, V., and Casabona, G. (2021). Measles, mumps, rubella prevention: How can we do better? *Expert Rev. Vaccines* 20 (7), 811–826. doi:10.1080/14760584.2021.1927722

Kazakova, E. V., Zghuang, T., Li, T., Fang, Q., Han, J., and Qiao, H. (2017). The Gas6 gene rs8191974 and Ap3s2 gene rs2028299 are associated with type 2 diabetes in the northern Chinese Han population. *Acta Biochim. Pol.* 64 (2), 227–231. doi:10.18388/abp.2016_1299

Kennedy, R. B., Ovsyannikova, I. G., Haralambieva, I. H., Lambert, N. D., Pankratz, V. S., and Poland, G. A. (2014). Genome-wide SNP associations with rubella-specific cytokine responses in measles-mumps-rubella vaccine recipients. *Immunogenetics* 66 (7-8), 493–499. doi:10.1007/s00251-014-0776-3

Liloglou, T., Walters, M., Maloney, P., Youngson, J., and Field, J. K. (2002). A T2517C polymorphism in the GSTM4 gene is associated with risk of developing lung cancer. *Lung Cancer* 37 (2), 143–146. doi:10.1016/s0169-5002(02)00078-8

Lin, H., Lunetta, K. L., Zhao, Q., Rong, J., Benjamin, E. J., Mendelson, M. M., et al. (2017). Transcriptome-wide association study of inflammatory biologic age. *Aging (Albany NY)* 9 (11), 2288–2301. doi:10.18632/aging.101321

Liu, D., Zhou, D., Sun, Y., Zhu, J., Ghoneim, D., Wu, C., et al. (2020). A transcriptome-wide association study identifies candidate susceptibility genes for pancreatic cancer risk. *Cancer Res.* 80 (20), 4346–4354. doi:10.1158/0008-5472.CAN-20-1353

Marshall, H. S., and Plotkin, S. The changing epidemiology of mumps in a high vaccination era. *Lancet Infect. Dis.* 2019;19(2):118–119. doi:10.1016/S1473-3099(18)30541-3

Masgras, I., Sanchez-Martin, C., Colombo, G., and Rasola, A. (2017). The chaperone TRAP1 as a modulator of the mitochondrial adaptations in cancer cells. *Front. Oncol.* 7, 58. . Published 2017 Mar 29. doi:10.3389/fonc.2017.00058

Oni, T. E., Biffi, G., Baker, L. A., Hao, Y., Tonelli, C., Somerville, T. D. D., et al. (2020). SOAT1 promotes mevalonate pathway dependency in pancreatic cancer. *J. Exp. Med.* 217 (9), e20192389. doi:10.1084/jem.20192389

Pridgeon, J. W., Olzmann, J. A., Chin, L. S., and Li, L. (2007). PINK1 protects against oxidative stress by phosphorylating mitochondrial chaperone TRAP1. *PLoS Biol.* 5 (7), e172. doi:10.1371/journal.pbio.0050172

Raghunathan, P. L., and Orenstein, W. (2022). Investing in global measles and rubella elimination is needed to avert deaths and advance health equity. *Lancet Glob. Health* 10 (10), e1363–e1364. doi:10.1016/S2214-109X(22)00388-6

Sabikunnahar, B., Lahue, K. G., Asarian, L., Fang, Q., McGill, M. M., Haynes, L., et al. (2022). Sex differences in susceptibility to influenza A virus infection depend on host genotype. *PLoS One* 17 (9), e0273050. Published 2022 Sep 16. doi:10.1371/journal.pone.0273050

Sanchez-Martin, C., Menon, D., Moroni, E., Ferraro, M., Masgras, I., Elsey, J., et al. (2021). Honokiol bis-dichloroacetate is a selective allosteric inhibitor of the mitochondrial chaperone TRAP1. *Antioxid. Redox Signal* 34 (7), 505–516. doi:10.1089/ars.2019.7972

Sanchez-Martin, C., Moroni, E., Ferraro, M., Laquatra, C., Cannino, G., Masgras, I., et al. (2020). Rational design of allosteric and selective inhibitors of the molecular chaperone TRAP1. *Cell Rep.* 31 (3), 107531. doi:10.1016/j.celrep.2020.107531

Scepanovic, P., Alanio, C., Hammer, C., Hodel, F., Bergstedt, J., Patin, E., et al. (2018). Human genetic variants and age are the strongest predictors of humoral immune responses to common pathogens and vaccines. *Genome Med.* 10 (1), 59. doi:10.1186/s13073-018-0568-8

Speed, D., Holmes, J., and Balding, D. J. (2020). Evaluating and improving heritability models using summary statistics. *Nat. Genet.* 52 (4), 458–462. doi:10.1038/s41588-020-0600-y

Thériault, S., Gaudreault, N., Lamontagne, M., Rosa, M., Boulanger, M. C., Messika-Zeitoun, D., et al. (2018). A transcriptome-wide association study identifies PALMD as a

susceptibility gene for calcific aortic valve stenosis. *Nat. Commun.* 9 (1), 988. . Published 2018 Mar 7. doi:10.1038/s41467-018-03260-6

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* 58 (1), 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x

Veturi, Y., and Ritchie, M. D. (2018). How powerful are summary-based methods for identifying expression-trait associations under different genetic architectures? *Pac Symp. Biocomput* 23, 228–239.

Voigt, E. A., Haralambieva, I. H., Larrabee, B. L., Kennedy, R. B., Ovsyannikova, I. G., Schaid, D. J., et al. (2018). Polymorphisms in the wilms tumor gene are associated with interindividual variations in rubella virus-specific cellular immunity after measles-mumps-rubella II vaccination. *J. Infect. Dis.* 217 (4), 560–566. doi:10.1093/infdis/jix538

Xiang, F., Ma, S. Y., Lv, Y. L., Zhang, D. X., Song, H. P., and Huang, Y. S. (2019). Tumor necrosis factor receptor-associated protein 1 regulates hypoxia-induced apoptosis through a mitochondria-dependent pathway mediated by cytochrome c oxidase subunit II. *Burns Trauma* 7, 16. 16. Published 2019 May 23. doi:10.1186/s41038-019-0154-3

Xiang, F., Ma, S. Y., Zhang, D. X., Zhang, Q., and Huang, Y. S. (2016). Tumor necrosis factor receptor-associated protein 1 improves hypoxia-impaired energy production in cardiomyocytes through increasing activity of cytochrome c oxidase subunit II. *Int. J. Biochem Cell Biol.* 79, 239–248. doi:10.1016/j.biocel.2016.08.041

Xiao, R., Gu, L., Li, A. M., Gan, Y. L., Liao, J. X., et al. (2023). IL-11 drives the phenotypic transformation of tracheal epithelial cells and fibroblasts to enhance abnormal repair after tracheal injury. *Biochim. Biophys. Acta Mol. Cell Res.* 1870 (4), 119438. doi:10.1016/j.bbamcr.2023.119438

Yang, M., Li, J. C., Tao, C., Wu, S., Liu, B., Shu, Q., et al. (2021). PAQR6 upregulation is associated with AR signaling and unfavorite prognosis in prostate cancers. *Biomolecules* 11 (9), 1383. . Published 2021 Sep 18. doi:10.3390/biom11091383

Yang, S., and Zhou, X. (2020). Accurate and scalable construction of polygenic scores in large biobank data sets. *Am. J. Hum. Genet.* 106 (5), 679–693. doi:10.1016/j.ajhg.2020.03.013

Yang, S., and Zhou, X. (2022). PGS-Server: Accuracy, robustness and transferability of polygenic score methods for biobank scale studies. *Brief. Bioinform* 23 (2), bbac039. doi:10.1093/bib/bbac039

Zhang, Y., Sun, Y., Zhu, H., Hong, H., Jiang, J., Yao, P., et al. (2022). Allelic imbalance of HLA-B expression in human lung cells infected with coronavirus and other respiratory viruses. *Eur. J. Hum. Genet.* 30 (8), 922–929. doi:10.1038/s41431-022-01070-5

Zheng, J., Zhao, Z., Ren, H., Wang, Y., Meng, X., Zhang, W., et al. (2022). LncRNA HCG11 facilitates nasopharyngeal carcinoma progression through regulating miRNA-490-3p/MAP3K9 Axis. *Front. Oncol.* 12, 872033. Published 2022 Apr 7, doi:10.3389/fonc.2022.872033

Zhou, X., Carbonetto, P., and Stephens, M. (2013). Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genet.* 9 (2), e1003264. doi:10.1371/journal.pgen.1003264

Zhu, T., Wang, Z., Zou, T., Xu, L., Zhang, S., Chen, Y., et al. (2021). SOAT1 promotes gastric cancer lymph node metastasis through lipid synthesis. *Front. Pharmacol.* 12, 769647. Published 2021 Nov 1, doi:10.3389/fphar.2021.769647

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 30 May 2023
DOI 10.3389/fgene.2023.1130529

Check for updates

OPEN ACCESS

EDITED BY
Yonghu Sun,
Shandong Provincial Hospital of Dermatology, China

REVIEWED BY
Koldo Garcia-Etxebarria,
Biodonostia Health Research Institute (IIS Biodonostia), Spain
Marina Laplana,
University of Lleida, Spain

*CORRESPONDENCE
Xiang Gao,
gxiang@mail.sysu.edu.cn
Donglin Ren,
rendl@mail.sysu.edu.cn

†These authors have contributed equally to this work

RECEIVED 23 December 2022
ACCEPTED 17 April 2023
PUBLISHED 30 May 2023

CITATION
Tang J, Wan X, Zhang J, Diao N, Zhang C, Gao X and Ren D (2023), A frameshift variant in the *SIRPB1* gene confers susceptibility to Crohn's disease in a Chinese population.
Front. Genet. 14:1130529.
doi: 10.3389/fgene.2023.1130529

COPYRIGHT
© 2023 Tang, Wan, Zhang, Diao, Zhang, Gao and Ren. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A frameshift variant in the *SIRPB1* gene confers susceptibility to Crohn's disease in a Chinese population

Jian Tang[1†], Xingyang Wan[2†], JunXiao Zhang[3†], Na Diao[1], Caibin Zhang[4], Xiang Gao[1*] and Donglin Ren[2*]

[1]Department of Gastroenterology, Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [2]Department of Colorectal and Anal Surgery, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [3]Institute of Biomedical Sciences, SequMed Biotech Inc., Guangzhou, China, [4]Institute of Clinical Pharmacology, School of Pharmaceutical Sciences, Sun Yat-Sen University, Guangzhou, China

Background: Crohn's disease (CD), a chronic gastrointestinal inflammatory disease, is increasing in China. With a focus on Han Chinese families with CD, the aim of this study was to find genetic variations that increase CD susceptibility by genome sequencing, genetic association, expression, and functional research.

Materials and methods: We performed family-based genome sequencing (WGS) analysis on 24 patients with CD from 12 families and then filtered shared potential causal variants by incorporating association results from meta-analyses of CD GWAS and immunology genes and *in silico* variant effect prediction algorithms. Replication analyses were performed in an independent cohort including 381 patients with CD and 381 control subjects.

Results: There were 92 genetic variants significantly associated with CD in Chinese individuals. Among them, 61 candidate loci were validated in replication analyses. As a result, patients carrying a rare frameshift variant (c.1143_1144insG; p. Leu381_Leu382fs) in gene *SIRPB1* had significantly higher risk to develop CD (p = 0.03, OR 4.59, 95% CI 0.98–21.36, 81.82% vs. 49.53%). The frameshift variation induced tyrosine phosphorylation of Syk, Akt, and Jak2, elevated the expression of *SIRPB1* at the mRNA and protein levels, activated DAP12, and controlled the activation of NF-κB in macrophages. Additionally, it promoted the synthesis of the pro-inflammatory cytokines IL-1, TNF-, and IL-6.

Conclusion: Our results suggest that the rare gain-of-function frameshift variant in *SIRPB1* is associated in Han Chinese patients with CD. The functional mechanism of *SIRPB1* and its downstream inflammatory pathways was preliminarily explored in CD.

KEYWORDS

Crohn 's disease, han Chinese patients, gene susceptibility, whole gene sequencing, SIRPB1

Introduction

Inflammatory bowel disease (IBD), which includes Crohn's disease (CD) and ulcerative colitis, is a complex polygenic disorder brought on by the improper activation of effector immunologic pathways in those with a genetic predisposition. Genome-wide association studies (GWAS) that look for genetic factors influencing disease start and progression have found 240 IBD-associated loci, which have greatly improved our understanding of the biology underlying these conditions (Anderson et al., 2011; de Lange et al., 2017; de Lange et al., 2017; Franke et al., 2010; Liu et al., 2015). The success of the GWAS approach is supported by the link between neighboring common variations in human populations, but it also makes it challenging to determine with accuracy which variant is causative, the molecular consequences of that variant, and frequently even which gene is perturbed. The association of uncommon protein-altering mutations that are expected to impart a higher risk of disease may help to explain some of the missing heritability in IBD. Because they are linked with fewer neighboring variants, rare variants with greater impact sizes may be easier to understand mechanistically. It is still unknown, nevertheless, how much of the heritability of complex disorders can be attributed to uncommon variations. The promise to better comprehend the molecular and genetic architecture of an exemplary complex disease is thus provided by well-powered investigations of uncommon variations in IBD.

IBD is assumed to have a significant genetic link because the most significant risk factor for the condition at any age is a family history of IBD (Childers et al., 2014; Kuwahara et al., 2012). Patients with IBD who have a family history of the condition frequently present it at a younger age, are more likely to have extra-intestinal manifestations, develop perforating disease, and need longer follow-up than patients without a family history, showing significant an increased genetic susceptibility to the condition (Kuwahara et al., 2012). Therefore, family-based IBD cohort genomic analyses are helpful for understanding the genetic architecture of IBD. IBD susceptibility is influenced by a variety of genetic variables. Whole genome sequencing (WGS), in conjunction with recent significant technological advancements, has made it possible to identify uncommon and novel harmful variants in IBD, providing a deeper understanding of genetic differences within the human genome. Due to this, GWAS have been created to objectively identify genetic risk factors for complex polygenic disorders. According to our hypothesis, rare or novel variations, such as those in genes linked to the innate immune system, are more likely to play a role in the development of CD in patients with a family history of the disease. We used WGS to investigate particular genes or pathways implicated in this disease state. Our capacity to research uncommon variations and ascertain the disease's genetic origin has been transformed by WGS. To address this question, in this study, we identified and replicated a novel rare causative frameshift variant associated with CD and explored into how it affected functionality.

Materials and methods

Study subjects

The genome sequence of 24 Han Chinese patients with CD from 12 families were obtained (Table 1). Written informed consent was provided by the attending parents or legal guardians of the pediatric participants. This study was approved by the Research Ethics Committee (REC) of the Sixth Affiliated Hospital of Sun Yet Sen University (SYSU) (N0.2020ZSLYEC-006).

WGS

Using a QIAamp DNA Kit (QIAGEN, Hilden, Germany), genomic DNA was extracted from peripheral blood cells in accordance with the manufacturer's protocol. DNA samples were quantified using a Qubit (Thermo Fisher Scientific). A total of 2 μg of each DNA sample was sent to the Beijing Genome Institute (BGI, Shenzhen, China) for WGS using the BGISEQ-500, according to the manufacturer's guidelines. According to the manufacturer's recommendations, the genomic DNA was briefly split by ultrasound on a Covaris E220 (Covaris) to DNA segments between 50 bp—800 bp. The fragmented DNA was then exposed to end-repair, phosphorylation, and A-tailing procedures after being further chosen to 100bp-300bp using AMpure XP Beads (Beckman Coulter, Indiana, United States). The A-tailed segments were ligated to the BGISEQ-500 platform-specific adaptors, and the ligated fragments were then purified and amplified using PCR. Finally, single-stranded DNA circles were produced by the circularization process. The libraries were sequenced using 50 bp paired-end reads on the BGISEQ-500 platform following quantification and qualifying.

Genome sequence data analysis

SOAPnuke was used to filter the raw sequencing reads (Li et al., 2009b) (N rating >10%, low quality rating >50%, and quality rating<5) and Burrows-Wheeler Aligner (BWA v0.7.17) to align to the UCSC human reference genome (hg19) (Li and Durbin, 2009; Li and Durbin, 2010). The coordinates were sorted using Samtools (version 1.3.1) and duplicates were identified using Picard (version 1.129, http://picard.sourceforge.net) (Li et al., 2009a). Using GATK HaplotypeCaller (McKenna et al., 2010), single nucleotide substitution variants (SNV) and brief insertions and deletions (indels) were identified (McKenna et al., 2010). We used the GATK Variant Quality Score Recalibration (VQSR) that uses machine learning algorithm to filter the raw variant callset. The GATK VQSR used high-quality known variant sets as training and truth resources and built a predictive model to filter spurious variants. The SNPs and InDels marked PASS in the output VCF file were high-confident variation set. For SNPs recalibration strategy, we used the following datasets and features to train the model. (a) Training sets: HapMap V3.3, Omni2.5 M genotyping array data and high-confidence SNP sites produced by the 1000 Genomes Project. (b) Features: Coverage (DP), Quality/depth (QD), Fisher

TABLE 1 The clinical characteristics of enrolled families of Crohn's disease.

Family number	Kinship	Age (years)	Sex	Disease duration (year)	Disease location	Disease behavior	Pelvic disease	Abdominal surgical history
F1-II-1	fraternal twins	18	male	5	L3	B3	yes	yes
F1-II-2		18	male	3	L3	B3	yes	no
F2-II-1	fraternal twins	22	female	7	L3+L4	B1	yes	no
F2-II-2		22	female	8	L3+L4	B1	yes	no
F3-I-3	aunt and nephew	42	female	1	L3	B3	no	yes
F3-II-1		19	male	<1	L3	B1	yes	no
F4-I-2	mother and daughter	45	female	10	L3	B3	yes	yes
F4-II-2		24	female	5	L3	B2	no	no
F5-I-1	father and son	44	male	18	L3	B2	no	yes
F5-II-1		21	male	<1	L3	B1	yes	no
F6-II-1	identical twins	14	female	<1	L3	B2	no	no
F6-II-2		14	Female	<1	L2	B2	yes	no
F7-II-1	elder sister and younger brother	23	female	2	L3	B1	no	no
F7-II-2		21	male	1	L3	B3	no	no
F8-I-1	father and son	44	male	2	L1+L4	B2	yes	no
F8-II-1		20	male	6	L1+L4	B3	yes	no
F9-II-1	elder and younger brother	31	male	1	L3+L4	B1	yes	no
F9-II-2		26	male	2	L3+L4	B1	yes	no
F10-I-1	father and daughter	45	male	7	L1	B2	no	yes
F10-II-1		14	female	5	L3	B3	no	no
F11-II-1	identical twins	37	male	11	L3	B3	no	yes
F11-II-2		37	male	11	L3	B3	no	no
F12-II-1	identical twins	15	female	1	L3	B1	yes	no
F12-II-2		15	female	1	L3	B1	no	no

Montreal classification' of Crohn's disease (CD); Disease location (L): L1 terminal ileum, L2 colon, L3 ileocolon, L4 upper gastrointestinal tract; Disease behavior (B): B1 non stricturing non penetrating; B2 stricturing, B3 penetrating.

test on strand bias (FS), Odds ratio for strand bias (SOR), Mapping quality rank sum test (MQRankSum), Read position rank sum test (ReadPosRankSum), RMS mapping quality (MQ). For InDels recalibration strategy, we used the following datasets and features to train the model. (a) Training sets: Mills 1000G gold standard InDel set. (b) Features: Coverage (DP), Quality/depth (QD), Fisher test on strand bias (FS), Odds ratio for strand bias (SOR), Mapping quality rank sum test (MQRankSum), Read position rank sum test (ReadPosRankSum).

Prioritization of variants

All germline SNV and indels were annotated using an in-house annotation pipeline, as described previously (Neveling et al., 2013; de Voer et al., 2013; Vissers et al., 2010). High-confidence calls (i.e., ≥10 reads, ≥5 variant reads, and ≥20% variant reads) were subsequently prioritized for variants that were non-synonymous and were absent in our in-house variant database (2,037 in-house analyzed exomes, mostly from European ancestry). Next, we removed all variants present with a MAF of >0.001 in dbSNPv138, the National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project database (ESP, 6503 exomes, http://evs.gs.washington.edu/EVS/), the Exome Aggregation Consortium (Lek et al., 2016) and Thousand Genome Project. Subsequently, family index patients shared non-synonymous variants that result in alterations in protein function, including protein truncation, splice site defects and missense mutations at highly conserved (phyloP ≥ 3.0) nucleotide positions, were included in our analyses. Alamut v. 2.0 software (Interactive Biosoftware) and integrated mutation prediction software (align GVDV, SIFT and PolyPhen-2) (Adzhubei et al., 2010; Kumar et al., 2009; Vissers et al., 2010) packages were used for analyses of the identified variants. The

prediction of splicing effects was evaluated based on five different algorithms (SpliceSiteFinder, MaxEntScan, NNSPLICE, GeneSplicer, Human Splicing Finder) through the bioinformatics tools of the Alamut v.2.0 software.

Identification and selection of variants of candidate genes

By using GWAS, we first targeted germline variations in genes at susceptibility loci known to be linked to IBD (Anderson et al., 2011; de Lange et al., 2017; Ellinghaus et al., 2016; Huang et al., 2017; Julia et al., 2014; Kenny et al., 2012; Liu et al., 2015; Parkes et al., 2007; Yamazaki et al., 2013; Yang et al., 2014). Genetic defects in innate immunity that impair intestinal bacterial sensing are linked to the development of IBD (Cananzi et al., 2021). Recent developments in molecular biology have uncovered crucial details about the genetic basis of numerous inflammatory diseases. Next to the identification of variants in known IBD GWAS genes, we searched for potential pathogenic variants in novel candidate genes using the remaining genome data of our CD family cohort. We concentrated on genes that fulfilled the following criteria while choosing these variants: (Supplementary Table S2): 1) genes with variations that caused protein truncation (such as putative frameshifts, nonsense variations, and variations at canonical splice sites), as well as non-synonymous variations with a PhyloP score of more than 3.0, were chosen; 2) the International Union of Immunological Societies Expert Committee on Primary Immunodeficiency's collection of primary immune deficiency (PID) genes (Picard et al., 2015); and 3) genes involved in pathways implicated in IBD pathogenesis, including innate immune system, immune system, and neutrophil degranulation pathway (Belinky et al., 2015; Kanegane, 2018).

Sanger sequencing

After PCR amplification, WGS-identified candidate variant of *SIRPB1* was verified using Sanger sequencing. The Primer3 software program was used to build PCR primers *in silico*. Standard PCR procedures were used on an Applied Biosystems Dual 96-Well GeneAmp PCR System 9,700 (primer sequences available upon request). Using the software package Vector NTI, variant analyses were carried out (Invitrogen, Paisley, United Kingdom).

Variant validation in independent cohort

Candidate variant validation analysis was performed on 381 probands with IBD and 381 unrelated individuals recruited by the Sixth Affiliated Hospital of SYSU using MassARRAY (Derkach et al., 2013). All individuals recruited were peotected by the REC of the Sixth Affiliated Hospital of SYSU (N0.2020ZSLYEC-006). The phenotypes of the controls were assessed, and no known gastrointestinal or immunological findings were reported. Fisher's exact testing was conducted, and statistical significance was set at $p < 0.05$ (Derkach et al., 2013).

Functional validation of the SIRPB1 variant allele

Histology

Both historical standard hematoxylin and eosin histological sections from the patient who was found to have the SIRPB1 p. Leu381 Leu382fs variation and IBD controls were assessed.

Expression analysis

On 4 μm segments of formalin-fixed paraffin-embedded (FFPE) tissue samples containing terminal ileum tissue from the patient identified as harboring the SIRPB1 p. Leu381 _Leu382fs variation and IBD controls, SIRPB1 expression was analyzed by immunohistochemistry (IHC). Using a BenchMark XT automated tissue staining machine (Ventana Medical Systems, Tucson, AZ, United States), IHC staining was carried out in accordance with the manufacturer's verified protocols.

Gene expression data

All the microarray samples used in this study were systematically searched and downloaded from NCBI-GEO (Barrett et al., 2013) after the manual curation of the sample details. We obtained gene expression data of mucosal biopsies in CD patients and normal controls from following array data series: GSE75214, GSE36807 and GSE59071. The bioinformatics online tool GEO2R was used to analyze the mRNA expression of SIRPB1.

Plasmid transfection

To study the variant type *in vitro* to understand its functional implications in more detail, we established THP-1 cell lines that stably expressed wild-type *SIRPB1* (*SIRPB1*^{wt}) and mutant *SIRPB1* (*SIRPB1*^{11143iG}, c.1143_1144insG; p. Leu381_Leu382fs). Human *SIRPB1*^{wt} and *SIRPB1*^{11143iG} expression vectors (pEZ-M02/*SIRPB1*^{wt} and pEZ-M02/*SIRPB1*^{11143iG}, respectively) were established. According to the manufacturer's instructions, Lipofectamine 2000 (Invitrogen; Thermo Fisher Scientific, Waltham, MA, United States) was used to transfect THP-1 cells with 0.5 μg of either pEZ-M02/*SIRPB1*^{wt} and pEZ-M02/*SIRPB1*^{11143iG}. Continuous neomycin treatment at 450 μg/mL was used to select transfectants that could consistently express the inserted vector plasmid (Thermo Fisher Scientific). The limited dilution approach was used to clone neomycin-resistant cells, which were then kept alive in medium containing neomycin.

Differentiation of THP-1 cells to macrophages

By administering THP-1 monocytes 100 ng/mL phorbol 12-myristate 13-acetate (PMA; Sigma) for 48 h, the macrophage-like state was generated. Then, THP-1 macrophages were transfected with 0.5 μg pEZ-M02/*SIRPB1*^{wt} and pEZ-M02/*SIRPB1*^{11143iG} plasmid for 4 h. After transfection, THP-1 macrophages were incubated in

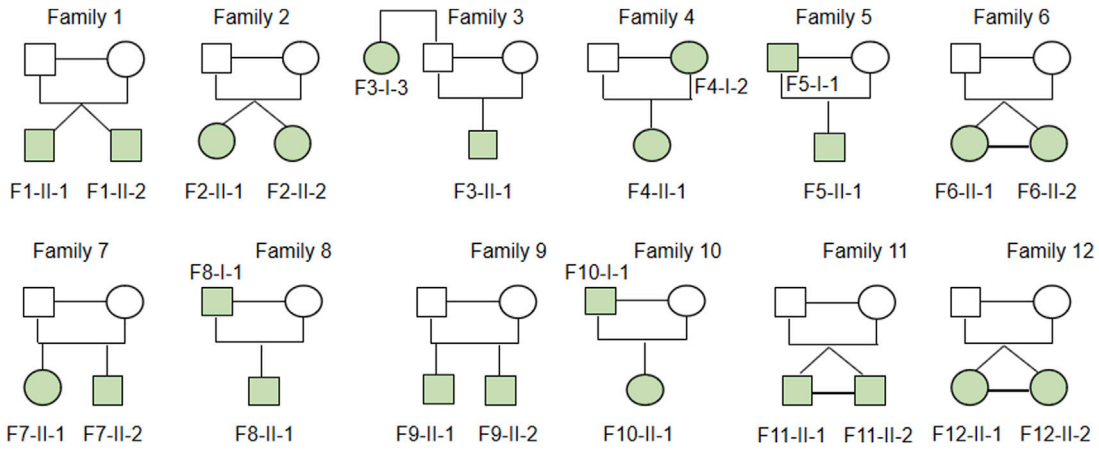


FIGURE 1
Heritability of enrolled CD family. Two pairs of fraternal twins, three pairs of identical twins, four pairs father or mother and daughter or son, two pairs of elder sister or brother and younger brother, one pair of aunt and niece were included in family based WGS analysis.

complete medium for 48 h and stimulated with LPS (100 ng/mL). THP-1 cells were lysed in ice-cold lysis buffer for western blotting, and proteins were separated on 10% SDS page. The main antibodies against DAP12, p-Syk/Syk, p-Akt/Akt, p-Jak2/Jak2, and anti-GAPDH were then used to probe the membranes. According to the manufacturer's instructions, the culture medium was suspended for ELISA assays and TNF-α, IL-1, and IL-6 ELISA kits were used to detect the substances.

Statistical analyses

Prism version 8.0 software (GraphPad) was used to perform the statistical analysis. Both the Student's t-test and the Dunnett's test were used to determine the significance of differences. When more than two groups were compared, a one-way ANOVA was performed. The data are shown as mean ± SE, and a significance level of 0.05 was used.

Results

Clinical characteristics of the enrolled patients with familial CD

Studies of twins and familial clustering of disease clearly indicate that IBD, especially CD, is a hereditary disorder (Halfvarson et al., 2003). Therefore, we hypothesized that family-based CD is more probable to have occurred as a result of uncommon or unique variations and may play a role in the disease's development. Thus, 24 patients with CD from 12 families in the case database of our IBD center were enrolled for WGS analysis. Table 1 provides a summary of the clinical traits of the individuals that were enrolled. Among them, there were two pairs of fraternal twins, three pairs of identical twins, four pairs of father or mother and daughter or son, two pairs of elder sister or brother and younger brother, and one pair of aunt and niece (Figure 1). All the patients were of Han origin, including 13 men (54.17%) and 11 women (45.83%). The median age of participants in this study was 22 years (range 14–45 years). Regarding disease location, 20 patients were diagnosed with ileocolonic inflammation (L3, 83.33%), one patient with colonic inflammation (L2, 4.17%), 3 patients with ileal inflammation (L1, 12.50%), and six patients had concomitant upper gastrointestinal disease (25.00%). Regarding disease behavior, nine patients were accompanied with non-stricturing non-penetrating lesions (B1, 37.50%), six patients had stricturing lesions (B2, 25.00%), and nine patients had penetrating lesions (B3, 37.50%). A total of 13 patients (54.17%) had pelvic disease and six patients (25.00%) had a history of abdominal surgery.

IBD susceptibility analysis of identified gene variant

To further understand the genetic basis of IBD, we used WGS to look for uncommon, possibly disease-causing coding variations in familiar CD patients (Table1). A substantial percentage of coding variants discovered across individuals in a family were prioritized for further exploration using a series of progressive variant filters. We identified on average 3,430,329 variants (range: 3,393,972–3,466,204) per genome. A prioritization scheme was applied to identify candidate variants as shown in table 2. After rigorous bioinformatics analysis steps (see Methods), there were 92 genetic variants significantly associated with CD in Chinese individuals including several known IBD related gene including *NOD2, ZPBP2, TNFSF15, LSP1, SDCCAG3, MUC19*. Then, we adopted the MassArray Analyzer system (Sequenom, Inc., SanDiego, CA, United States) for genotyping those genetic variants. We designed primers for 89 variants in Sequenom official websites, while, there is suitable primer for the remaining 3 variants. However, those 89 primers were distributed in 5 panels, due to the limited budget, only the top 2 largest panels (a total of 61 variants were included, about 30 variants per panel) used for the subsequently assay. We finally generated a candidate variants list

TABLE 2 Prioritization scheme for genome data analysis of 24 CD patients.

Type of prioritization filter	Remaining variants (*n*)
All variants	82,327,899
Coding region and canonical splice site variants after quality filtering (total ≥ 10 reads, ≥ 5 variant reads and ≥ 20% variant reads)	13,819
Non-synonymous variants, canonical splice site variants	9,833
Family index patients sharedVariants that result in alterations in protein function (protein truncation, splice site defects and missense mutations at highly conserved (phyloP ≥ 3.0) nucleotide positions.	2,749
Not in in-house database and MAF ≤ 0.001 in dbSNPv138, ESP, ExAC and Thousand Genome Project	2,007
Variants in genes at susceptibility loci known to be linked to IBD, primary immune deficiencies gene and IBD-related pathway genes and novel candidate variants	132
Variants pass IGV check	92
Variants for validation in large cohort	61 (4 varints in GWAS genes, 22 variants in IBD related pathway genes and 35 novel variants)

including 61 variants for replication analyses with an independent cohort including 381 patients with CD and 381 control subjects (Table 3).

Gain-of-function frameshift variant SIRPB1 p. Leu381_Leu382fs in a cohort of patients with CD and healthy controls

Notably, we identified the frameshift variant *SIRPB1* p. Leu381_Leu382fs in two index patients, F7-II-1 and F7-II-2, from Family 7 (Figure 2). Patient F7-II-1 was the elder sister of patient F7-II-2, whose age was 23 years with a disease duration of about 2 years, and patient F7-II-2 was the younger brother whose age was 21 years with a disease duration of approximately 1 year. Both manifested as ileocolonic inflammation (L3). The disease behavior of F7-II-1 was non-stricturing non-penetrating (B1), whereas the disease behavior of F7-II-2 was penetrating (B3). For replication analysis, the blood sample of 384 CD patients including 104 females and 280 males in our database were used for microarray analysis to verify the susceptibility gene mutation. The mean age of these patients was 27.0 ± 10.4 years-old. The 384 blood samples of control group were collected from healthy testing population including 170 female and 214 female and the mean age of these patients was 34.7 ± 9.1 years-old. The unique uncommon frameshift variant in *SIRPB1* was considerably enriched in CD patients compared to controls, according to genotyping analysis (p = 0.03, OR 4.59, 95% CI 0.98–21.36, 81.82% vs. 49.53%). (Table 3). The clinical characteristics of patients with CD identified to harbor the SIRPB1 p. Leu381_Leu382fs variant are listed in Table 4.

IHC analysis of the SIRPB1 p. Leu381_Leu382fs variant

A member of the family of signal-regulating proteins (SIRP) and of the immunoglobulin superfamily, SIRPβ (also known as CD172b) is encoded by *SIRPB1* (van den Berg et al., 2005). Previous research on *SIRPB1* mostly focused on its biochemical properties and functions and discovered that it stimulates DAP12 and Syk tyrosine phosphorylation, which then activates the mitogen-activated protein kinase (MAPK) pathway to increase phagocytosis in macrophages (Hayashi et al., 2004). However, the role of *SIRPB1* in IBD pathogenesis has not yet been reported. Therefore, we first assessed the expression levels of *SIRPB1* by analyzing publicly available data from patients with IBD and healthy individuals. According to the findings, patients with active CD (A-CD) had considerably higher levels of *SIRPB1* expression in their ileocolonic tissue than either healthy individuals or patients with CD that was in remission (R-CD) (Figures 3G–I), indicating that *SIRPB1* is possibly involved in the progression of CD. Next, we assessed the expression levels of SIRPβ in the ileocolonic tissue of variant and wild-type patients with CD using IHC (Supplementary Table S3). Notably, patients with CD and the *SIRPB1* variant exhibited significantly higher levels of SIRPβ than those with CD and wild-type *SIRPB1* (Figures 3A,D). Consistently, SIRPβ-mediated transduction signal molecules, such as DAP12 and p-Syk, were also significantly upregulated in patients with CD and the variant compared with WT controls (Figures 3B,C,E,F). Our data suggest that the higher expression of SIRPβ was caused by the frameshift variation of *SIRPB1*.

Functional characterization of the rare SIRPB1 frameshift variant (c.1143_1144insG; p. Leu381_Leu382fs)

As documented above, the frameshift variant in *SIRPB1* led to higher expression of SIRPβ; however, the function of the rare *SIRPB1* frameshift variant still needs to be elucidated. The interaction of SIRPβ with the activating adaptor protein DAP12, which carries an immunoreceptor tyrosine-based activation motif (ITAM) and transmits activating signals, depends on the presence of

TABLE 3 IBD susceptibility analysis of identified gene variants.

Gene	*rs number*	Unique ID	Genotype (mutation vs. wild type)	Genotype frequency (mutation carries vs. wild type carries)	P value [a]	OR	95% CI
PCNXL2	rs759917992	chr1:233386580T>C	TT[b]	NA[c]	NA[c]	NA[c]	NA[c]
TP73	rs1641123267	chr1:3638706C>T	CT vs. CC[d]	**Patients group:** 1/379(0.26%) vs. 378/379(99.74%) **Control group:** 0/365(0.00%) vs. 365/365(100.00%)	1.000	1.00	1.00-1.01
UBE2U	rs776567811	chr1:64676467AC>A	AC[b]	NA[c]	NA[c]	NA[c]	NA[c]
LSP1	NA	chr11:1891892G>A	CC[b]	NA[c]	NA[c]	NA[c]	NA[c]
C11orf42	rs751350529	chr11:6231170C>T	CT vs. CC[d]	**Patients group:** 0/379(0.00%) vs. 379/379(100.00%) **Control group:** 2/380(0.53%) vs. 378/380(99.47%)	0.499	1.00	0.99-1.00
C11orf42	rs749213397	chr11:6231682G>A	GA vs. GG[d]	**Patients group:** 0/380(0.00%) vs. 380/380(100.00%) **Control group:** 3/374(0.80%) vs. 371/374(99.20%)	0.122	0.99	0.98-1.00
ALG8	rs1488580557	chr11:77813960G>A	GG[b]	NA[c]	NA[c]	NA[c]	NA[c]
POLR3B	rs2036678883	chr12:106770153C>T	CC[b]	NA[c]	NA[c]	NA[c]	NA[c]
OAS3	rs771052891	chr12:113385843GC>G	GC.G vs. GC.GC[d]	**Patients group:** 6/380(1.58%) vs. 374/380(98.42%) **Control group:** 2/370(0.54%) vs. 368/370(99.46%)	0.287	2.95	0.59-14.72
MUC19	rs1411441474	chr12:40834995T>A	TA vs. TT[d]	**Patients group:** 2/380(0.79%) vs. 378/380(99.47%) **Control group:** 6/372(1.61%) vs. 366/372(98.39%)	0.173	0.32	0.07-1.61
MYO1A	NA	chr12:57440644TAC>T	TAC.TAC[b]	NA[c]	NA[c]	NA[c]	NA[c]
A2ML1	rs766100204	chr12:8998099G>A	GA vs. GG[d]	**Patients group:** 2/374(0.53%) vs. 372/374(99.47%) **Control group:** 1/363(0.28%) vs. 362/363(99.72%)	1.000	1.95	0.18-21.56
EPSTI1	NA	chr13:43537471T>TA	TT[b]	NA[c]	NA[c]	NA[c]	NA[c]

(Continued on following page)

TABLE 3 (*Continued*) IBD susceptibility analysis of identified gene variants.

Gene	*rs number*	Unique ID	Genotype (mutation vs. wild type)	Genotype frequency (mutation carries vs. wild type carries)	P value [a]	OR	95% CI
DAAM1	NA	chr14:59782026G>T	GT vs. GG[d]	**Patients group:**	1.000	1.00	1.00-1.01
				1/378(0.26%) vs. 377/378(99.74%)			
				Control group:			
				0/364(0.00%) vs. 364/364(100.00%)			
SYNE2	rs777169796	chr14:64519635A>T	AT vs. AA[d]	**Patients group:**	1.000	1.00	1.00-1.01
				1/378(0.26%) vs. 377/378(99.74%)			
				Control group:			
				0/354(0.00%) vs. 354/354(100.00%)			
CPPED1	rs748886359	chr16:12798820C>T	CC[b]	NA[c]	NA[c]	NA[c]	NA[c]
BEAN1	rs989514270	chr16:66471600C>A	CC[b]	NA[c]	NA[c]	NA[c]	NA[c]
AATF	rs1306922955	chr17:35307665G>A	GG[b]	NA[c]	NA[c]	NA[c]	NA[c]
ZPBP2	rs1460554471	chr17:38024800A>G	AA[b]	NA[c]	NA[c]	NA[c]	NA[c]
PDK2	rs748085033	chr17:48185985C>T	CT + TT vs. CC	**Patients group:**	1.000	1.01	0.06-16.13
				1/380(0.26%) vs. 379/380(99.74%)			
				Control group:			
				1/382(0.26%) vs. 381/382(99.74%)			
ENO3	rs764120380	chr17:4856098C>T	CC[b]	NA[c]	NA[c]	NA[c]	NA[c]
EPX	rs757233476	chr17:56281773C>T	CT vs. CC[d]	**Patients group:**	0.499	1.00	1.00-1.01
				1/379(0.26%) vs. 378/379(99.74%)			
				Control group:			
				0/381(0.00%) vs. 381/381(100.00%)			
COL5A3	NA	chr19:10079057	TCACAGGGTCTCC.T vs. TCACAGGGTCTCC.TCACAGGGTCTCC[d]	**Patients group:**	0.499	1.00	1.00-1.01
				1/379(0.26%) vs. 378/379(99.74%)			
		TCACAGGGTCTCC>T		**Control group:**			
				0/381(0.00%) vs. 381/381(100.00%)			
TYK2	NA	**chr19:10479075**	**GAAGC.G + GG** vs. **GAAGC.GAAGC**	**Patients group:**	**0.006**	**0.10**	**0.01-0.76**
				1/380(0.26%) vs. 379/380(99.74%)			
		GAAGC>G		**Control group:**			
				10/375(2.67%) vs. 365/375(97.33%)			

(Continued on following page)

TABLE 3 (*Continued*) IBD susceptibility analysis of identified gene variants.

Gene	rs number	Unique ID	Genotype (mutation vs. wild type)	Genotype frequency (mutation carries vs. wild type carries)	P value [a]	OR	95% CI
GDF1	rs1568291627	chr19:18981025 ACGGGGGCG>A	ACGGGGGCG.A + AA vs. ACGGGGGCG.ACGGGGGCG	**Patients group:** 1/381(0.26%) vs. 380/381(99.74%) **Control group:** 1/364(0.27%) vs. 363/364(99.73%)	1.000	0.96	0.06-15.33
C19orf40	rs760353712	chr19:33464372C>CTT	C. CTT vs. CC[d]	**Patients group:** 1/377(0.27%) vs. 376/377(99.73%) **Control group:** 0/378(0.00%) vs. 378/373787(100.00%)	0.499	1.00	1.00-1.01
VRK3	rs1048569809	chr19:50528523C>G	CG vs. CC[d]	**Patients group:** 4/380(1.05%) vs. 376/380(98.95%) **Control group:** 1/380(0.26%) vs. 379/380(99.74%)	0.373	4.03	0.45-36.24
DFNB59	rs1437628682	chr2:179320735A>G	AG vs. AA[d]	**Patients group:** 1/67(1.49%) vs. 66/67(98.51%) **Control group:** 0/103(0.00%) vs. 103/103(100.00%)	0.394	1.02	0.99-1.05
VRK2	rs1328945383	chr2:58312086G>A	GG[b]	NA[c]	NA[c]	NA[c]	NA[c]
WDPCP	NA	chr2:63486522T>TC	T.TC vs. TT[d]	**Patients group:** 1/380(0.26%) vs. 379/380(99.74%) **Control group:** 0/345(0.00%) vs. 345/345(100.00%)	1.000	1.00	1.00-1.01
ANKEF1	rs752349062	chr20:10019057C>A	CC[b]	NA[c]	NA[c]	NA[c]	NA[c]
SIRPB1	rs1275744950	chr20:1546854GC>GCC	GC.GCC vs. GC[d]	**Patients group:** 9/381(2.36%) vs. 372/381(97.64%) **Control group:** 2/381(0.52%) vs. 379/381(99.48%)	**0.034**	**4.59**	**0.98-21.36**
CPNE1	NA	chr20:34214629C>CT	C.CT vs. CC[d]	**Patients group:** 1/380(0.26%) vs. 379/380(99.74%) **Control group:** 0/380(0.00%) vs. 380/380(100.00%)	1.000	1.00	1.00-1.01

(Continued on following page)

TABLE 3 (*Continued*) IBD susceptibility analysis of identified gene variants.

Gene	*rs number*	Unique ID	Genotype (mutation vs. wild type)	Genotype frequency (mutation carries vs. wild type carries)	P value [a]	OR	95% CI
PLCG1	NA	chr20:39792446A>T	AA[b]	NA[c]	NA[c]	NA[c]	NA[c]
FBXO40	NA	chr3:121341344CT>C	CT.C vs. CT.CT[d]	**Patients group:**	1.000	1.00	1.00-1.01
				1/379(0.26%) vs. 378/379(99.74%)			
				Control group:			
				0/379(0.00%) vs. 379/379(100.00%)			
ERC2	rs1293707325	chr3:56183136G>A	GG[b]	NA[c]	NA[c]	NA[c]	NA[c]
MMAA	rs757548934	chr4:146572222C>T	CT vs. CC[d]	**Patients group:**	1.000	1.00	1.00-1.01
				1/380(0.26%) vs. 379/380(99.74%)			
				Control group:			
				0/339(0.00%) vs. 339/339(100.00%)			
KLHL5	rs755006031	chr4:39105132G>GT	GG[b]	NA[c]	NA[c]	NA[c]	NA[c]
SEC31A	NA	chr4:83788384G>A	GG[b]	NA[c]	NA[c]	NA[c]	NA[c]
SLC2A9	NA	chr4:9922067C>T	CT vs. CC[d]	**Patients group:**	1.000	1.00	1.00-1.00
				0/380(0.00%) vs. 380/380(100.00%)			
				Control group:			
				1/382(0.26%) vs. 381/382(99.74%)			
FBXL21P	rs201662172	chr5:135273232C>A	CA vs. CC[d]	**Patients group:**	1.000	1.00	1.00-1.01
				1/379(0.26%) vs. 378/379(99.74%)			
				Control group:			
				0/379(0.00%) vs. 379/379(100.00%)			
PCDH12	NA	**chr5:141336148C>T**	**CT vs. CC[d]**	**Patients group:**	**0.006**	**0.10**	**0.01-0.77**
				1/380(0.26%) vs. 379/380(99.74%)			
				Control group:			
				10/380(2.63%) vs. 370/380(97.37%)			
GHR	rs752025877	chr5:42565977A>G	AA[b]	NA[c]	NA[c]	NA[c]	NA[c]
SYNJ2	NA	chr6:158438246	AAAGG[b]	NA[c]	NA[c]	NA[c]	NA[c]
		AAAGG>A					
CAGE1	rs1414911763	**chr6:7329418G>A**	**GA vs. GG[d]**	**Patients group:**	**1.9×10^{-5}**	**0.90**	**0.86-0.94**
				0/174(0.00%) vs. 174/174(100.00%)			
				Control group:			
				23/232(9.91%) vs. 209/232(90.09%)			

(Continued on following page)

TABLE 3 (*Continued*) IBD susceptibility analysis of identified gene variants.

Gene	*rs number*	Unique ID	Genotype (mutation vs. wild type)	Genotype frequency (mutation carries vs. wild type carries)	P value [a]	OR	95% CI
SH2B2	NA	chr7:101960938C>T	CC[b]	NA[c]	NA[c]	NA[c]	NA[c]
RPA3	rs529874466	chr7:7758145C>G	CG vs. CC[d]	**Patients group:**	1.000	0.84	0.05-13.55
				1/373(0.27%) vs. 372/373(99.73%)			
				Control group:			
				1/315(0.32%) vs. 314/315(99.68%)			
CD36	rs748202229	chr7:80285946C>T	CC[b]	NA[c]	NA[c]	NA[c]	NA[c]
TNFSF15	NA	chr9:117568285T>C	TC + CC vs. TT	**Patients group:**	0.499	1.01	1.00-1.01
				2/381(0.52%) vs. 379/381(99.48%)			
				Control group:			
				0/381(0.00%) vs. 381/381(100.00%)			
LRSAM1	rs770106776	chr9:130265135G>A	GA + AA vs. GG	**Patients group:**	0.622	0.49	0.04-5.44
				1/379(0.26%) vs. 378/379(99.74%)			
				Control group:			
				2/373(0.54%) vs. 371/373(99.46%)			
HMCN2	rs1445146226	chr9:133245203G>A	GA vs. GG[d]	**Patients group:**	0.499	1.00	1.00-1.01
				1/379(0.26%) vs. 378/379(99.74%)			
				Control group:			
				0/381(0.00%) vs. 381/381(100.00%)			
SDCCAG3	rs375609278	chr9:139301649C>G	CG vs. CC[d]	**Patients group:**	0.490	1.00	0.99-1.00
				0/380(0.00%) vs. 380/380(100.00%)			
				Control group:			
				1/365(0.27%) vs. 364/365(99.73%)			
PRSS3	NA	chr9:33794797TGA>T	TGA[b]	NA[c]	NA[c]	NA[c]	NA[c]
CCL27	rs746707552	chr9:34662369G>A	GG[b]	NA[c]	NA[c]	NA[c]	NA[c]
PCSK5	rs769457551	chr9:78973443C>T	CT vs. CC[d]	**Patients group:**	1.000	1.00	0.06-16.05
				1/380(0.26%) vs. 379/380(99.74%)			
				Control group:			
				1/380(0.26%) vs. 379/380(99.74%)			
NOD2	rs104895438	chr16:50745656G>A	GG[b]	NA[c]	NA[c]	NA[c]	NA[c]
MUC19	**rs112524759**	**chr12:40882387TA>T**	**TA.T + TT vs. TA.TA**	**Patients group:**	**0.03**	**1.01**	**0.73-1.40**

(Continued on following page)

TABLE 3 (*Continued*) IBD susceptibility analysis of identified gene variants.

Gene	*rs number*	Unique ID	Genotype (mutation vs. wild type)	Genotype frequency (mutation carries vs. wild type carries)	P value [a]	OR	95% CI
				97/379(25.59%) vs. 282/379(74.41%)			
				Control group:			
				94/370(25.41%) vs. 276/370(74.59%)			
BIRC8	rs145690856	chr19:53793456G>A	GG[b]	NA[c]	NA[c]	NA[c]	NA[c]
XPA	rs149226993	chr9:100447247G>A	GG[b]	NA[c]	NA[c]	NA[c]	NA[c]
GLB1	rs192732174	chr3:33109737G>A	GG[b]	NA[c]	NA[c]	NA[c]	NA[c]
ERCC4	rs2020959	chr16:14041622C>A	CA vs. CC[d]	**Patients group:**	1.00	1.00	1.00-1.01
				1/377(0.27%) vs. 376/377(99.73%)			
				Control group:			
				0/363(0.00%) vs. 363/363(100.00%)			

[a]Chi-Square Tests or Fisher's Exact Test;
[b]Only one genotype was detected;
[c]Not available, because this mutation/SNP have no minor allele;
[d]Only two genotypes were detected; These P-values ≤0.05 were highlighted in bold font.

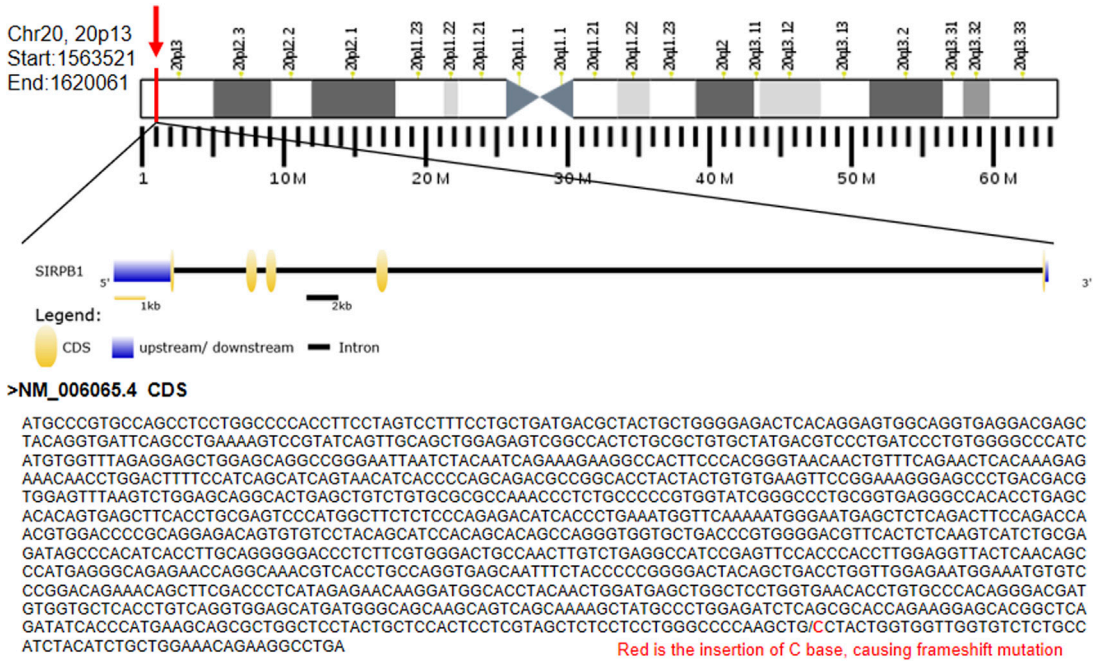


FIGURE 2
Identification of a rare frameshift mutation of *SIRPB1* gene A. rare frameshift mutation (c.1143_1144insG; p. Leu381_Leu382fs) was identified in an innate immune gene SIRPB1 in both probands of Family CD-7.

a basic amino acid side chain in the transmembrane domain of SIRPβ (Lanier and Bakker, 2000; Tomasello and Vivier, 2005). An earlier study showed that phosphorylating Syk and MAPK and crosslinking mouse SIRPβ with monoclonal antibodies increases neutrophil migration and macrophage phagocytosis (Hayashi et al., 2004; Liu et al., 2005). As shown in Figure 4, the frameshift variant of

TABLE 4 The clinical characteristics of CD patients with *SIRPB1* gene mutation.

Patient number	Age (years)	sex	Disease duration (year)	Disease location	Disease behaviour	Pelvic disease	Abdominal surgical history
1	23	female	2	L3	B1	yes	no
2	21	male	1	L3	B3	no	no
3	20	male	<1	L3	B1	no	no
4	17	male	1	L3	B1	yes	no
5	42	male	8	L3	B3	no	yes
6	15	female	5	L3	B1	yes	no
7	31	male	6	L3	B3	no	yes
8	21	male	2	L3	B1	no	no
9	54	male	20	L1	B1	no	no

Montreal classification' of Crohn's disease (CD); Disease location (L): L1 terminal ileum, L2 colon, L3 ileocolon, L4 upper gastrointestinal tract; Disease behavior (B): B1 non stricturing non penetrating; B2 stricturing, B3 penetrating.

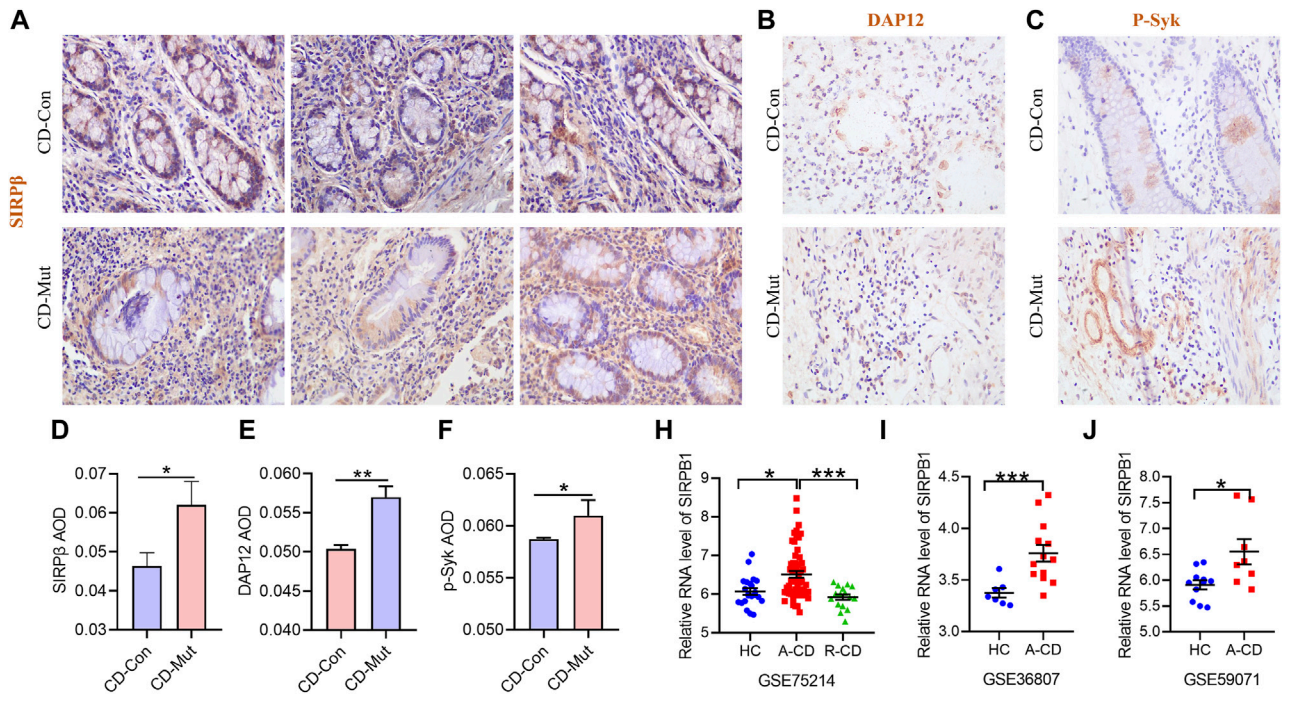


FIGURE 3
Higher expression of *SIRPB1* in mutant CD patients. A/B/C Immunohistochemistry (IHC) staining of SIRPβ **(A)**, DAP12 **(B)**, and p-Syk **(C)** in ileocolonic tissue from wild type CD patients and control patients; **(D,E,F)** Statistical analyses analysis of IHC staining; G/H/I *SIRPB1* expression was assessed using a publicly accessible data. The following array data series were analyzed to generate the human patient expression data: GSE75214; GSE36807; GSE59071. GSE75214 **(G)**: mucosal biopsies of ileal mucosa of 11 health controls (HC), 15 inactive CD patients (R-CD) and 50 active CD patients (A-CD) were used to analyze mRNA expression of SIRPB1. GSE36807 **(H)**: intestinal biopsies from 7 health controls (HC), and 15 CD patients were used to analyze mRNA expression of SIRPB1. GSE59071 **(I)** mucosal biopsies were obtained at endoscopy from 8 health controls (HC) and Crohn's disease (CD) patients.

SIRPB1 gene led to an alteration of the amino acid sequence located both on the transmembrane and the cytoplasm. Therefore, we hypothesized that the frameshift variant of *SIRPB1* could make functional contributions to the inflammation process.

To investigate the potential effects of the identified novel frameshift variant of *SIRPB1*, Plasmids carrying either $SIRPB1^{wt}$ and $SIRPB1^{11143iG}$ sequence were transfected into THP-1 cells that had been stimulated with PMA for 48 h. In line with a previous

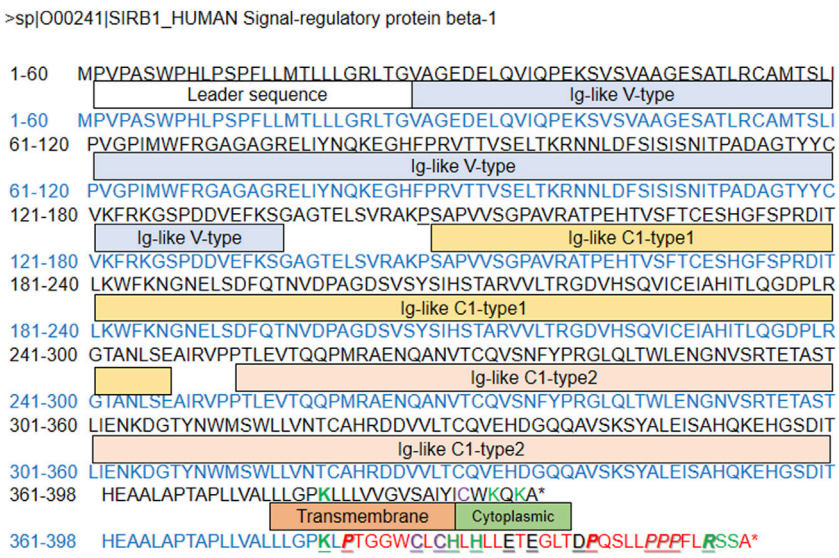


FIGURE 4
Predicted amino-acid sequence from wild-type control and patients with frameshift SIRPB1 mutation.

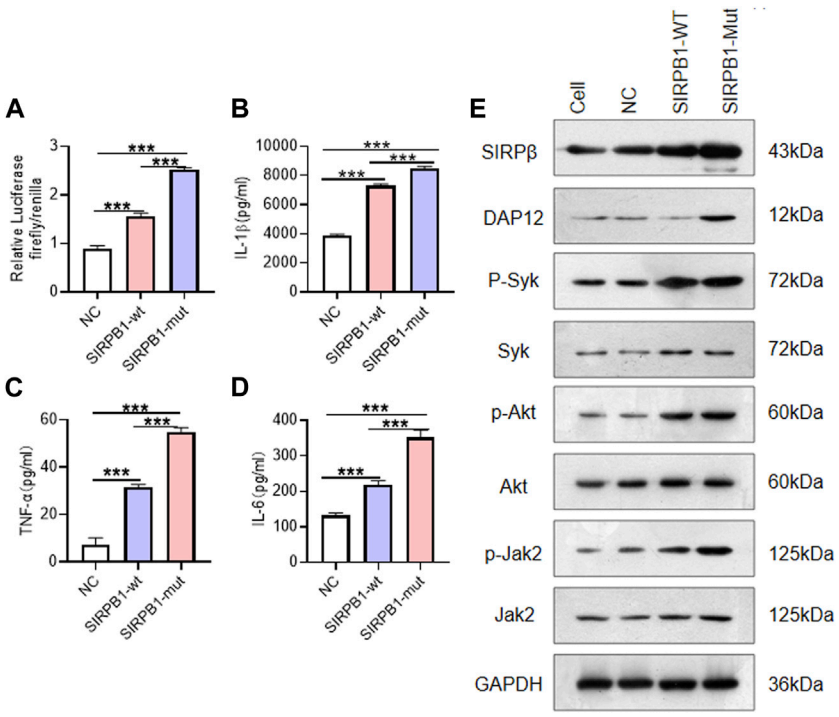


FIGURE 5
SIRPB1 mutation promote inflammatory response of macrophages *in vitro*. **(A)** NF-kB transcriptional activity of THP-1 cells from different groups was detected by a luciferase reporter assay; **(B,C,D)** IL-1β, IL-6 and TNFα levels in culture medium of THP-1 cells from different groups; **(E)** Western Blotting of signal transduction molecules in THP-1 cells from different groups.

report, compared with THP-1 cells transfected with the empty vector (NC), THP-1 cells with *SIRPB1* overexpression ($SIRPB1^{wt}$) exhibited increased activation of DAP12 and NF-κB and elicited tyrosine phosphorylation of Syk, Akt, and Jak2, causing increased IL-1β, TNF-α, and IL-6 release (Figures 5A,B,C,D,E). Furthermore, compared with $SIRPB1^{wt}$ cells, THP-1 cells expressing variant *SIRPB1* ($SIRPB1^{11143iG}$) displayed higher SIRPβ expression, enhanced activation of DAP12 and NF-κB, increased tyrosine

phosphorylation of Syk, AKT, and Jak2, and higher secretion of IL-1β, TNF-α, and IL-6 (Figures 5A,B,C,D,E). Taken together, our results show that the frameshift variant of *SIRPB1* amplified the function of the wild-type *SIRPB1* gene.

Discussion

SIRPs are a class of cell surface signaling receptors that are produced differently in myeloid and neural cells. They each include three extracellular Ig-like domains (Adams et al., 1998; Kharitonenkov et al., 1997). Despite having very identical extracellular domains, SIRPs can be distinguished as activating (α) or inhibitory (β) isoforms using conventional patterns in their cytoplasmic or transmembrane regions (Kharitonenkov et al., 1997). SIRPα, which binds to its ligand CD47, was found to inhibit signaling pathways that are mediated by receptor tyrosine kinase, as its cytoplasmic domains contain immunoreceptor tyrosine-based inhibitory motifs (ITIMs), which recruit the phosphatase SH2-domain-containing proteins SHP-1 and SHP-2 *in vivo* (Barclay and Brown, 2006). Contrary to SIRPα, SIRPβ lacks the sequence patterns necessary to attach SHP-1 and SHP-2 in its short six-amino acid cytoplasmic domain. The transmembrane domain of SIRPβ, however, has a charged amino acid residue that can bind to DAP12, which possesses a single cytoplasmic ITAM, and activate cell-mediated cytotoxicity and cytokine release (Dietrich et al., 2000; Liu et al., 2005; McVicar et al., 1998).

In the present study, a rare frameshift variant (c.1143_1144insG; p. Leu381_Leu382fs) in the innate immunity gene *SIRPB1* was identified. Genotyping analysis revealed that the novel frameshift variant in *SIRPB1* was significantly enriched in patients compared to that in healthy controls. Based on publicly available gene expression data, our research showed that patients with CD in the active stage had considerably higher relative expression of SIRPβ, indicating that SIRPβ is involved in the progression of intestinal inflammation. In addition, the frameshift variant of *SIRPB1* leads to an alteration of the amino acid sequence located both in the transmembrane and cytoplasmic regions. Therefore, we tested the functional contribution of the frameshift variant of *SIRPB1*. Our data show that, compared with WT controls, THP-1 cells transfected with the variant *SIRPB1* sequence displayed higher expression of SIRPβ and its adaptor protein DAP12, enhanced activation of subsequent signal transduction, and increased pro-inflammatory cytokines production and NF-κB expression. Accordingly, IHC data of ileocolonic tissue implied that patients with CD and variant *SIRPB1* expressed higher levels of SIRPβ and its adaptor protein DAP12. Collectively, we found a rare frameshift variant (c.1143_1144insG; p. Leu381_Leu382fs) in *SIRPB1* that leads to alterations in the amino acid sequence located both in the transmembrane and cytoplasmic domains, could contribute to inflammation exacerbation by promoting the expression of SIRPβ and its adaptor protein, DAP12. We hypothesized that the frameshift variant in *SIRPB1* may result in conformational changes of SIRPβ or form a functional cytoplasmic domain owing to a longer amino acid sequence, which needs to be elucidated in future studies.

In our study, the data analysis shows that the rare gain-of-function frameshift variant (c.1143_1144insG; p. Leu381_Leu382fs) in *SIRPB1* is associated with Han Chinese patients with CD and provides insights that the variant in *SIRPB1* upregulated the activation of NF-κB and increased the production of IL-1β, TNF-α, and IL-6 by inducing DAP12, Syk, Akt, and Jak2 to become tyrosine phosphorylated in macrophages, CD pathogenesis is facilitated. This study has several limitations. The familiar CD patients and the subsequent cohort for replication analyses are all Han Chinese from a single IBD center, so, further validation in other population is necessary. As well, due to limited budget and lack of suitable primers, only 61 variants were validated in replication analyses. Although the function of frameshift variant in SIRPB1 were explored *in vitro* using THP-1 cells, further investigation using other cell lines and also *in vivo* are required. And also, for replication analysis, a p-value <0.05 without multiple testing correction was used in validation, so, further function study was adopted to verify the susceptive possibility of the SIRPB1 variant in CD patients.

Data availability statement

The data presented in the study are deposited in the SRA repository and can be accessed at: https://dataview.ncbi.nlm.nih.gov/object/PRJNA962047.

Ethics statement

The studies involving human participants were reviewed and approved by Research Ethics Committee (REC) of the Sixth Affiliated Hospital of Sun Yet Sen University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

XG and DR as correspondent author: conception and design, revising the article, final approval of the version to be published. JT, XW, and JZ as co-first author: conception and design, drafting the article, final approval of the version to be published. ND: analysis and interpretation of data, drafting the article, final approval of the version to be published. CZ: analysis and interpretation of data, final approval of the version to be published.

Funding

This project was supported by the National Nature Science Foundation of China (81870382), the Major Scientific and Technological Project of Guangdong Province (A2020123), the Science and Technology Program of Guangzhou (201903010081), and the Youth Project for National Nature Science Foundation of China (82100547).

Acknowledgments

We thank the patients and their families contributing to this study.

Conflict of interest

Author JZ was empolyed by SequMed Biotech Inc.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1130529/full#supplementary-material

References

Adams, S., van der Laan, L. J., Vernon-Wilson, E., Renardel, D. L. C., Dopp, E. A., Dijkstra, C. D., et al. (1998). Signal-regulatory protein is selectively expressed by myeloid and neuronal cells. *J. Immunol.* 161 (4), 1853–1859. doi:10.4049/jimmunol.161.4.1853

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods.* 7 (4), 248–249. doi:10.1038/nmeth0410-248

Anderson, C. A., Boucher, G., Lees, C. W., Franke, A., D'Amato, M., Taylor, K. D., et al. (2011). Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat. Genet.* 43 (3), 246–252. doi:10.1038/ng.764

Barclay, A. N., and Brown, M. H. (2006). The SIRP family of receptors and immune regulation. *Nat. Rev. Immunol.* 6 (6), 457–464. doi:10.1038/nri1859

Belinky, F., Nativ, N., Stelzer, G., Zimmerman, S., Iny, S. T., Safran, M., et al. (2015). PathCards: Multi-source consolidation of human biological pathways. *Database (Oxford)* 2015, 2015. doi:10.1093/database/bav006

Cananzi, M., Wohler, E., Marzollo, A., Colavito, D., You, J., Jing, H., et al. (2021). IFIH1 loss-of-function variants contribute to very early-onset inflammatory bowel disease. *Hum. Genet.* 140 (9), 1299–1312. doi:10.1007/s00439-021-02300-4

Childers, R. E., Eluri, S., Vazquez, C., Weise, R. M., Bayless, T. M., and Hutfless, S. (2014). Family history of inflammatory bowel disease among patients with ulcerative colitis: A systematic review and meta-analysis. *J. Crohns Colitis.* 8 (11), 1480–1497. doi:10.1016/j.crohns.2014.05.008

de Lange, K. M., Moutsianas, L., Lee, J. C., Lamb, C. A., Luo, Y., Kennedy, N. A., et al. (2017). Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* 49 (2), 256–261. doi:10.1038/ng.3760

de Voer, R. M., Geurts, V. K. A., Weren, R. D., Ligtenberg, M. J., Smeets, D., Fu, L., et al. (2013). Germline mutations in the spindle assembly checkpoint genes BUB1 and BUB3 are risk factors for colorectal cancer. *Gastroenterology* 145 (3), 544–547. doi:10.1053/j.gastro.2013.06.001

Derkach, A., Lawless, J. F., and Sun, L. (2013). Robust and powerful tests for rare variants using Fisher's method to combine evidence of association from two or more complementary tests. *Genet. Epidemiol.* 37 (1), 110–121. doi:10.1002/gepi.21689

Dietrich, J., Cella, M., Seiffert, M., Buhring, H. J., and Colonna, M. (2000). Cutting edge: Signal-regulatory protein beta 1 is a DAP12-associated activating receptor expressed in myeloid cells. *J. Immunol.* 164 (1), 9–12. doi:10.4049/jimmunol.164.1.9

Ellinghaus, D., Jostins, L., Spain, S. L., Cortes, A., Bethune, J., Han, B., et al. (2016). Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat. Genet.* 48 (5), 510–518. doi:10.1038/ng.3528

Franke, A., McGovern, D. P., Barrett, J. C., Wang, K., Radford-Smith, G. L., Ahmad, T., et al. (2010). Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat. Genet.* 42 (12), 1118–1125. doi:10.1038/ng.717

Halfvarson, J., Bodin, L., Tysk, C., Lindberg, E., and Jarnerot, G. (2003). Inflammatory bowel disease in a Swedish twin cohort: A long-term follow-up of concordance and clinical characteristics. *Gastroenterology* 124, 1767–1773. doi:10.1016/s0016-5085(03)00385-8

Hayashi, A., Ohnishi, H., Okazawa, H., Nakazawa, S., Ikeda, H., Motegi, S., et al. (2004). Positive regulation of phagocytosis by SIRPbeta and its signaling mechanism in macrophages. *J. Biol. Chem.* 279, 29450–29460. doi:10.1074/jbc.M400950200

Huang, H., Fang, M., Jostins, L., Umicevic, M. M., Boucher, G., Anderson, C. A., et al. (2017). Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature* 547 (7662), 173–178. doi:10.1038/nature22969

Julia, A., Domenech, E., Chaparro, M., Garcia-Sanchez, V., Gomollon, F., Panes, J., et al. (2014). A genome-wide association study identifies a novel locus at 6q22.1 associated with ulcerative colitis. *Hum. Mol. Genet.* 23 (25), 6927–6934. doi:10.1093/hmg/ddu398

Kanegane, H. (2018). Inflammatory bowel diseases and primary immunodeficiency diseases. *Immunol. Med.* 41 (4), 154–161. doi:10.1080/25785826.2018.1556025

Kenny, E. E., Pe'Er, I., Karban, A., Ozelius, L., Mitchell, A. A., Ng, S. M., et al. (2012). A genome-wide scan of Ashkenazi Jewish Crohn's disease suggests novel susceptibility loci. *PLoS Genet.* 8 (3), e1002559. doi:10.1371/journal.pgen.1002559

Kharitonenkov, A., Chen, Z., Sures, I., Wang, H., Schilling, J., and Ullrich, A. (1997). A family of proteins that inhibit signalling through tyrosine kinase receptors. *Nature* 386 (6621), 181–186. doi:10.1038/386181a0

Kumar, P., Henikoff, S., and Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4 (7), 1073–1081. doi:10.1038/nprot.2009.86

Kuwahara, E., Asakura, K., Nishiwaki, Y., Inoue, N., Watanabe, M., Hibi, T., et al. (2012). Effects of family history on inflammatory bowel disease characteristics in Japanese patients. *J. Gastroenterol.* 47 (9), 961–968. doi:10.1007/s00535-012-0558-3

Lanier, L. L., and Bakker, A. B. (2000). The ITAM-bearing transmembrane adaptor DAP12 in lymphoid and myeloid cell function. *Immunol. Today* 21 (12), 611–614. doi:10.1016/s0167-5699(00)01745-x

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536 (7616), 285–291. doi:10.1038/nature19057

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25 (14), 1754–1760. doi:10.1093/bioinformatics/btp324

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26 (5), 589–595. doi:10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009a). The sequence alignment/map format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp352

Li, R., Yu, C., Li, Y., Lam, T. W., Yiu, S. M., Kristiansen, K., et al. (2009b). SOAP2: An improved ultrafast tool for short read alignment. *Bioinformatics* 25 (15), 1966–1967. doi:10.1093/bioinformatics/btp336

Liu, Y., Soto, I., Tong, Q., Chin, A., Buhring, H. J., Wu, T., et al. (2005). SIRPbeta1 is expressed as a disulfide-linked homodimer in leukocytes and positively regulates neutrophil transepithelial migration. *J. Biol. Chem.* 280 (43), 36132–36140. doi:10.1074/jbc.M506419200

Liu, J. Z., van Sommeren, S., Huang, H., Ng, S. C., Alberts, R., Takahashi, A., et al. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* 47 (9), 979–986. doi:10.1038/ng.3359

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20 (9), 1297–1303. doi:10.1101/gr.107524.110

McVicar, D. W., Taylor, L. S., Gosselin, P., Willette-Brown, J., Mikhael, A. I., Geahlen, R. L., et al. (1998). DAP12-mediated signal transduction in natural killer cells. A dominant role for the Syk protein-tyrosine kinase. *J. Biol. Chem.* 273 (49), 32934–32942. doi:10.1074/jbc.273.49.32934

Neveling, K., Feenstra, I., Gilissen, C., Hoefsloot, L. H., Kamsteeg, E. J., Mensenkamp, A. R., et al. (2013). A post-hoc comparison of the utility of sanger sequencing and exome

sequencing for the diagnosis of heterogeneous diseases. *Hum. Mutat.* 34 (12), 1721–1726. doi:10.1002/humu.22450

Parkes, M., Barrett, J. C., Prescott, N. J., Tremelling, M., Anderson, C. A., Fisher, S. A., et al. (2007). Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat. Genet.* 39 (7), 830–832. doi:10.1038/ng2061

Picard, C., Al-Herz, W., Bousfiha, A., Casanova, J. L., Chatila, T., Conley, M. E., et al. (2015). Primary immunodeficiency diseases: An update on the classification from the international union of immunological Societies Expert committee for primary immunodeficiency 2015. *J. Clin. Immunol.* 35 (8), 696–726. doi:10.1007/s10875-015-0201-1

Tomasello, E., and Vivier, E. (2005). KARAP/DAP12/TYROBP: Three names and a multiplicity of biological functions. *Eur. J. Immunol.* 35 (6), 1670–1677. doi:10.1002/eji.200425932

van den Berg, T. K., van Beek, E. M., Buhring, H. J., Colonna, M., Hamaguchi, M., Howard, C. J., et al. (2005). A nomenclature for signal regulatory protein family members. *J. Immunol.* 175 (12), 7788–7789. doi:10.4049/jimmunol.175.12.7788

Vissers, L. E., de Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., de Vries, P., et al. (2010). A de novo paradigm for mental retardation. *Nat. Genet.* 42 (12), 1109–1112. doi:10.1038/ng.712

Yamazaki, K., Umeno, J., Takahashi, A., Hirano, A., Johnson, T. A., Kumasaka, N., et al. (2013). A genome-wide association study identifies 2 susceptibility Loci for Crohn's disease in a Japanese population. *Gastroenterology* 144 (4), 781–788. doi:10.1053/j.gastro.2012.12.021

Yang, S. K., Hong, M., Zhao, W., Jung, Y., Baek, J., Tayebi, N., et al. (2014). Genome-wide association study of Crohn's disease in Koreans revealed three new susceptibility loci and common attributes of genetic susceptibility across ethnic populations. *Gut.* 63(1), 80–87. doi:10.1136/gutjnl-2013-305193

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 31 May 2023
DOI 10.3389/fgene.2023.1141010

Check for updates

OPEN ACCESS

EDITED BY
Yonghu Sun,
Shandong Provincial Hospital of Dermatology, China

REVIEWED BY
Hongsheng Gui,
Henry Ford Health System, United States
Sheng Yang,
Nanjing Medical University, China

*CORRESPONDENCE
Wilson Liao,
wilson.liao@ucsf.edu

RECEIVED 09 January 2023
ACCEPTED 09 May 2023
PUBLISHED 31 May 2023

CITATION
Bui A, Kumar S, Liu J, Orcales F, Gulliver S, Tsoi LC, Gulliver W and Liao W (2023), A partitioned 88-loci psoriasis genetic risk score reveals HLA and non-HLA contributions to clinical phenotypes in a Newfoundland psoriasis cohort.
Front. Genet. 14:1141010.
doi: 10.3389/fgene.2023.1141010

COPYRIGHT
© 2023 Bui, Kumar, Liu, Orcales, Gulliver, Tsoi, Gulliver and Liao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A partitioned 88-loci psoriasis genetic risk score reveals HLA and non-HLA contributions to clinical phenotypes in a Newfoundland psoriasis cohort

Audrey Bui[1,2], Sugandh Kumar[1], Jared Liu[1], Faye Orcales[1], Susanne Gulliver[3], Lam C. Tsoi[4,5,6], Wayne Gulliver[3,7] and Wilson Liao[1]*

[1]Department of Dermatology, University of California San Francisco, San Francisco, CA, United States, [2]Lake Erie College of Osteopathic Medicine, Bradenton, FL, United States, [3]NewLab Clinical Research Inc, St. John's, NL, Canada, [4]Department of Dermatology, University of Michigan, Ann Arbor, MI, United States, [5]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, United States, [6]Department of Biostatistics, University of Michigan, Ann Arbor, MI, United States, [7]Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, Canada

Psoriasis is an immune-mediated inflammatory skin disease typically characterized by erythematous and scaly plaques. It affects 3% of the Newfoundland population while only affecting 1.7% of the general Canadian population. Recent genome-wide association studies (GWAS) in psoriasis have identified more than 63 genetic susceptibility loci that individually have modest effects. Prior studies have shown that a genetic risk score (GRS) combining multiple loci can improve psoriasis disease prediction. However, these prior GRS studies have not fully explored the association of GRS with patient clinical characteristics. In this study, we calculated three types of GRS: one using all known GWAS SNPs (GRS-ALL), one using a subset of SNPs from the HLA region (GRS-HLA), and the last using non-HLA SNPs (GRS-noHLA). We examined the relationship between these GRS and a number of psoriasis features within a well characterized Newfoundland psoriasis cohort. We found that both GRS-ALL and GRS-HLA were significantly associated with early age of psoriasis onset, psoriasis severity, first presentation of psoriasis at the elbow or knee, and the total number of body locations affected, while only GRS-ALL was associated with a positive family history of psoriasis. GRS-noHLA was uniquely associated with genital psoriasis. These findings clarify the relationship of the HLA and non-HLA components of GRS with important clinical features of psoriasis.

KEYWORDS

genetic risk score (GRS), polygenic risk score (PRS), psoriasis, Newfoundland and labrador, genetics, HLA

1 Introduction

Psoriasis is an immune-mediated chronic inflammatory disease characterized by erythematous and scaly skin plaques. It affects 3% of the Newfoundland population while only affecting 1.7% of the general Canadian population (Nall et al., 1999; Papp et al., 2011). The Newfoundland province of Canada is a genetically isolated population

comprising approximately 500,000 residents with 98% English or Irish descent (Rahman et al., 2003). Because Newfoundland rose from a limited founder population, it is an exceptional resource for studying familial disorders such as psoriasis which was found at a higher prevalence in Newfoundland compared to other white populations (Nall et al., 1999).

Although environmental triggers like stress, infection, and trauma can contribute to the development of psoriasis, family-based and population studies suggest an important genetic component to the development of psoriasis (Alshobaili et al., 2010; Chandran and Raychaudhuri, 2010; Capon, 2017). The pathogenesis of psoriasis involves the dysregulation of T cells, antigen presenting cells, and keratinocytes among many cell types. Genome-wide association studies (GWAS) in psoriasis have identified many susceptibility loci with modest individual effects (Cargill et al., 2007; Stuart et al., 2015; Tsoi et al., 2017). Among these, the *HLA-C*06:02* allele is thought to contribute the greatest genetic effect (Henseler and Christophers, 1985). A prior study showed that with each additional *HLA-C*06:02* risk allele (tagged by rs10484554), there was a 206% elevated risk of psoriasis (Chen et al., 2011). However, *HLA-C*06:02* only accounted for 6.7% of the genetic heritability found in psoriasis, suggesting the importance of other HLA genes and non-HLA genes in disease progression (Chen et al., 2011).

While psoriasis commonly presents on the scalp and extremities, it often first occurs on the extensor surfaces such as the elbows and knees. Psoriasis is classified into two types: Type 1 psoriasis (T1P) which starts before age 40, and T2P which starts at or after the age of 40 (Henseler and Christophers, 1985). T1P has been associated with a family history of psoriasis with involvement of *HLA-C*06:02* whereas T2P is less associated with a positive family history or involvement of *HLA-C*06:02*.

Prior studies have shown that combining many genetic susceptibility loci into an overall genetic risk score (GRS) can improve identification of people at risk for that disease (Weedon et al., 2006; Meigs et al., 2008; Abraham et al., 2021). Previous GRS studies in psoriasis have shown that GRS can be used for psoriasis disease prediction and shows an inverse correlation with age of onset (Chen et al., 2011; Lu et al., 2013; Tsoi et al., 2017). However, there is a lack of studies investigating the association of GRS with psoriasis clinical features.

In this study, we calculated three types of GRS: one using all known GWAS SNPs (GRS-ALL), one using a subset of SNPs from the HLA region (GRS-HLA), and the last using non-HLA SNPs (GRS-noHLA). In past GWAS studies of psoriasis, *HLA-C* has been identified as a highly associated locus. Our lab and others have uncovered many functional roles for HLA genes, so it is intuitive to set these SNPs apart as a functionally separate group of SNPs as they play a major role in antigen presentation (Chen et al., 2012; Zeng et al., 2013; Yanovsky et al., 2020; Ahn et al., 2021). The partitioning of non-HLA SNPs was performed to investigate their possible role independently from the HLA SNPs. We examined the relationship between these GRS and several psoriasis features within a Newfoundland cohort. These include family history, age of onset, psoriasis severity, locations ever affected by psoriasis, locations first affected by psoriasis, and total number of locations. In addition, we explore how the association of GRS with these clinical features is modified by early-onset psoriasis *versus* late-onset psoriasis (T1P vs T2P).

2 Materials and methods

2.1 Cohort

The study cohort includes 654 psoriasis cases of European ancestry from the Newfoundland region of Canada. Psoriasis patients were residents of Newfoundland and Labrador that were clinically diagnosed with psoriasis by a dermatologist in the late 1980s and early 1990s. The cohort consisted of 52% female and 48% male patients with and average age of 41 with a standard deviation (SD) of 14. Clinical data collected included birthplace, maternal heritage, paternal heritage, race, BMI, gender, age of onset, relatives with psoriasis, body locations ever or first noted with psoriasis (hand, back of hand, palm, foot, sole, toe, nail, scalp, face, neck, arm, elbow, armpit, leg, knee, back, chest, trunk, genital area, other), other serious illnesses, psoriasis severity as rated by the dermatologist (i.e., mild, moderate, severe), and type of psoriasis (i.e., plaques, guttate, pustular, *etc.*). All data fields and category options are included in Supplementary Table S1. All subjects provided written informed consent for use of their data and biosamples under IRB approval HREB #2019.188.

The independent validation cohort data were collected from 345 psoriasis patients of European descent at the University of California San Francisco (UCSF) Department of Dermatology enrolled between 2006 and 2016. The diagnosis of psoriasis was confirmed by a board-certified dermatologist, and study subjects completed a survey including demographic characteristics, medical history, and clinical features. The cohort consisted of 48% female and 52% male patients with an average age of 49 and a SD of 16. All subjects provided written informed consent for use of their data and biosamples under UCSF IRB# 10-02830.

DNA from both cohorts was genotyped on the Affymetrix United Kingdom Biobank Axiom Array (ThermoFisher) using a GeneTitan Multi-Channel Instrument (Applied Biosystems). SNPs were called using Analysis Power Tools 2.10.2.2 (Affymetrix, https://www.affymetrix.com/support/developer/powertools/changelog/index.html). Sample and SNPs passed quality control using the parameters of call rate >97% and Dish QC >82%.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 (http://pngu.mgh.harvard.edu/purcell/plink/) (Purcell et al., 2007), and the remaining sites were sorted using Plink (www.cog-genomics.org/plink/2.0/) (Chang et al., 2015). Clinical study data were managed using REDCap (Research Electronic Data Capture) hosted at University of California San Francisco (Harris et al., 2009; Harris et al., 2019). This web-based software provides an intuitive interface for validated data capture, audit trails for tracking data manipulation and exportation,

automated export procedures, and procedures for data integration with external sources.

2.2 SNP selection

We constructed our GRS using 88 SNPs and their corresponding odds ratios (ORs) from the largest published psoriasis GWAS meta-analysis identifying SNPS meeting genome-wide significance ($p <$ 5.0E-08) (Supplementary Table S2) (Tsoi et al., 2017). This meta-analysis included six GWAS, one exomechip, and one immunochip datasets of European ancestry. Many of the psoriasis loci from this study contained secondary independent signals. To ensure the effect sizes from the primary signals were not over- or underestimated, the ORs were calculated by conditioning on other independent signals within the same locus (Tsoi et al., 2017). This ensured the independent effect of each signal on psoriasis was properly represented despite any linkage disequilibrium (LD) structure. Moreover, because of the especially high LD with HLA loci, we calculated the pairwise linkage disequilibrium between the 11 HLA SNPs across a European population using the LDmatrix Tool within LDlink (Machiela and Chanock, 2015). This confirmed low LD between the 11 HLA SNPs selected, with 93% of the pairwise LD comparisons having an $R^2 < 0.1$.

2.3 Imputation

SNP data from the Affymetrix United Kingdom Biobank array were augmented with imputed SNPs from 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/). The imputation quality of the 88 psoriasis SNPs was highly accurate with $R^2 > 0.8$ showing the high confidence and correct imputation of all SNPs (Supplementary Table S3). We utilized all 88 markers in our analysis as previous studies have indicated that utilizing imputation quality cutoffs has a detrimental impact on GRS discriminatory ability (Goldstein et al., 2015; Chen et al., 2020). Data is publicly available at https://doi.org/10.6084/m9.figshare.21970847.

2.4 GRS calculation

Initially, two approaches were used to calculate the GRS: the simple risk allele count (cGRS and the weighted method (GRS). Subsequent analyses found that the weighted GRS to be superior to cGRS, so only the weighted GRS results are presented. The weighted GRS was calculated as the sum of the number of risk alleles weighted by the OR of that allele. GRS calculation was performed on imputed continuous-valued dosages between 0 and 2 rather than number of risk alleles. The GRS-ALL was calculated as the weighted GRS using all 88 SNPs, the GRS-HLA was calculated using the 11 SNPs found in the HLA region (chr6:28510120-33480577 in GRCh38), and the GRS-noHLA was calculated using the 77 non-HLA SNPs. The calculations were performed in Plink.

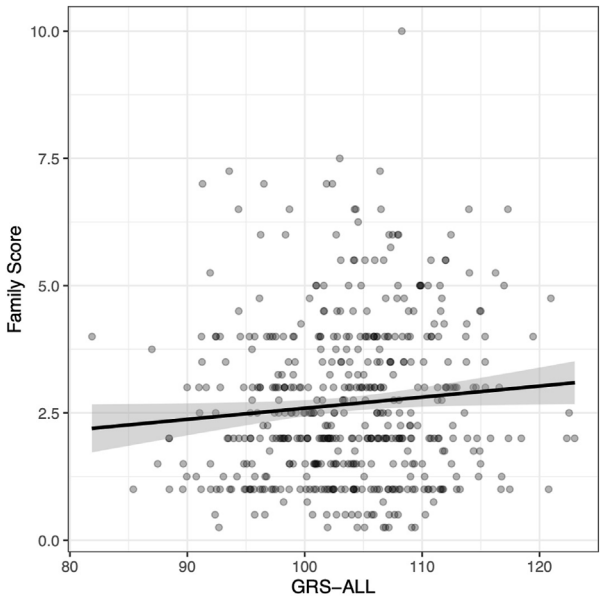


FIGURE 1
Positive correlation between family score and GRS-ALL. Coefficient = 0.022. $R^2 = 0.0067$, $p = 0.038$, SE = 0.011, n = 495. Family score calculated by summing the contributions from relatives affected by psoriasis as follows: first degree relative = 1, a second-degree relative = 0.5, a third-degree relative = 0.25.

2.5 Association testing

Association testing of GRS-ALL, GRS-HLA, and GRS-noHLA with psoriasis clinical features was performed using logistic regression or linear regression in R (v4.1.0). A p-value of less than 0.05 was deemed significant. The principal component analysis for genital psoriasis was performed in R (v4.1.0) using the prcomp () function in the stats (v4.1.0) package. The samples were grouped based on the presence of genital psoriasis. For each group, the top ten non-HLA SNPs in principal component 1 with the loadings of the greatest magnitude were further explored. Associations between the 88 loci and genital psoriasis were tested with SNPTEST v2.5.4 using the additive model ("-frequentist 1") and genotype dosages imputed by MIS ("-method expected") for all cohorts (Supplementary Table S4) (Wellcome Trust Case Control, 2007).

3 Results

3.1 Family history

The family score was calculated as the sum of all relatives diagnosed with psoriasis in which first-degree relatives were given a weight of 1, second-degree relatives were given a weight of 0.5, and third-degree relatives were given a weight of 0.25. A linear regression was used to evaluate the relationship between GRS and psoriasis family score. A significant positive association was found between GRS-ALL and the family score ($p = 0.038$; Figure 1) with and odds ratio (OR) [95% CI] of 1.02 [1.00-1.04]. Meanwhile, this

association was found to be nonsignificant when using GRS-HLA and GRS-noHLA. These results indicate that all SNPs regardless of loci cumulatively contribute to the family history.

3.2 Age of psoriasis onset

A linear regression analysis revealed a significant negative association between GRS-ALL and onset in which increasing GRS was associated with an earlier age of onset (coefficient = −0.341, standard error (SE) = 0.09, R^2 = 0.023, p = 0.0001, OR 0.71 [0.59-0.85]). The average age of onset was calculated for each GRS-ALL score quartile in which the first quartile had an average age of onset of 27.28 (SD = 13.76, SE = 1.07) while the fourth quartile had an average of 20.63 (SD = 12.43, SE = 0.97; Figure 2; Supplementary Table S5). The linear regression was repeated using GRS-HLA which yielded a more significant association (coefficient = −0.772, SE = 0.156, R^2 = 0.0398, $p = 1.03 \times 10^{-6}$, OR 0.46 [0.34-0.63]). Meanwhile, the results showed a similar trend but were nonsignificant when repeated with the GRS-noHLA. These results suggest that individuals with a higher GRS have an earlier age of onset and that this trend is driven by the HLA SNPs.

3.3 Psoriasis severity

An ordinal logistic regression was performed to evaluate the relationship between GRS and psoriasis severity level, categorized as mild, moderate, and severe, as rated by a dermatologist. The patients were categorized based on dermatologist diagnosis and discretion. There was a significant association between GRS-ALL and severity (SE = 0.0152, p = 0.045, OR 1.02 [1.00-1.06]) and between GRS-HLA and severity (SE = 0.027, $p = 1.64 \times 10^{-3}$, OR 1.09 [1.03-1.15]; Figure 3A; Supplementary Table S6). Meanwhile, there was a non-significant association between GRS-noHLA and severity. To ensure these results were not driven by non-plaque forms of psoriasis (e.g., pustular, erythrodermic) being considered more severe, a sensitivity analysis was performed that included only plaque psoriasis (n = 361), which also confirmed the association of GRS-ALL and GRS-HLA with severity while there was no significant association with GRS-noHLA. These results suggest that the HLA SNPs are primarily responsible for psoriasis severity.

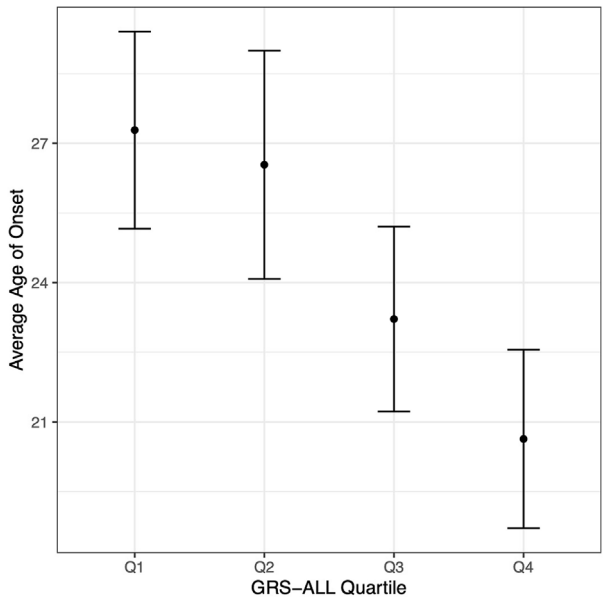


FIGURE 2
Age of onset decreases with increasing GRS-ALL. Point corresponds to average age of onset. Vertical bars correspond to 95% confidence intervals. n = 163 per quartile.

3.4 Locations ever affected by psoriasis

Linear regression revealed that GRS-ALL correlated positively with the total number of locations ever affected by psoriasis (p = 1.25e-10, SE = 0.023, n = 649, OR 1.16 [1.11-1.21]; Supplementary Figure S1), with a stronger positive correlation observed between GRS-HLA and total locations ever affected (coefficient = 0.295, R^2 = 0.077, $p = 3.37 \times 10^{-13}$, SE = 0.040, n = 649, OR 1.34 [1.24-1.45]). Meanwhile, there was a nonsignificant weaker positive correlation when using GRS-noHLA.

A logistic regression was also performed to evaluate the effects of GRS on appearance of psoriasis at 30 different locations (Supplementary Table S7). GRS-ALL and GRS-HLA were both significantly associated with a number of individual body locations. With respect to the incidence of psoriasis at specific locations, only GRS-noHLA had a significant effect on the appearance of psoriasis in the genital area, which was not observed in GRS-ALL and GRS-HLA. These results indicate that patients with higher GRS have more total body locations ever noted with psoriasis, many of which are driven by HLA SNPs; however, the genital area was strongly associated with non-HLA SNPs.

Intrigued by this novel association of genital psoriasis with non-HLA SNPs, we examined this in an independent UCSF cohort of 345 European psoriasis patients of which 32% have genital psoriasis (110 genital, 235 non-genital). In this cohort, genital psoriasis was also found to be strongly associated with GRS-noHLA and GRS-ALL ($p < 0.05$) while it was not associated with GRS-HLA (Supplementary Table S7). This independent dataset provides additional evidence for the association between genital psoriasis and non-HLA SNPs.

To explore the biological basis for the association between non-HLA SNPs and genital psoriasis, an association test between the 88 loci and genital psoriasis was performed. Of these, five SNPs and four SNPs were found to be significant in the Newfoundland and UCSF cohort, respectively, but were found to be insignificant after false discovery rate (FDR) adjustment (Supplementary Table S4). A principal component analysis was also performed to identify which non-HLA SNPs contribute the greatest discriminatory power for genital *versus* non-genital psoriasis (Supplementary Figure S2). SNPs identified as contributing to genital psoriasis included those near *IFIH1* (rs3747517), *UBE2L3* (rs2256609), rs12651787, and rs100040411. SNPs identified as contributing to non-genital psoriasis included *IFIH1* (rs2111485), *ERAP1* (rs39841), *DDX58* (rs11795343), rs2057338, and rs8128234.

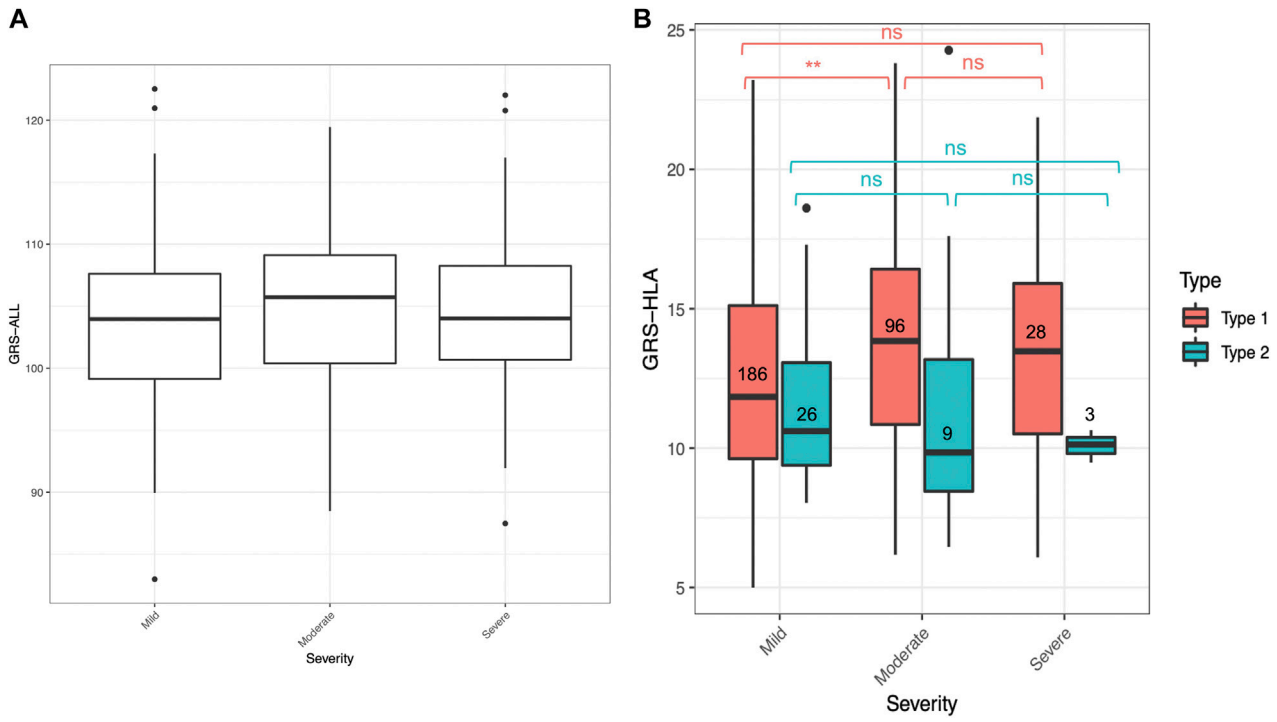


FIGURE 3
GRS for mild, moderate, and severe psoriasis. Horizontal lines represent quartiles. Points represent outliers. **(A)** Non-significant difference in GRS-ALL between severity levels when analyzed with Kruskal–Wallis test. n = 265 (mild), 112 (moderate), 32 (severe). **(B)** Significant difference in GRS-HLA between mild and moderate in Type 1 psoriasis. Kruskal–Wallis Test *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$.

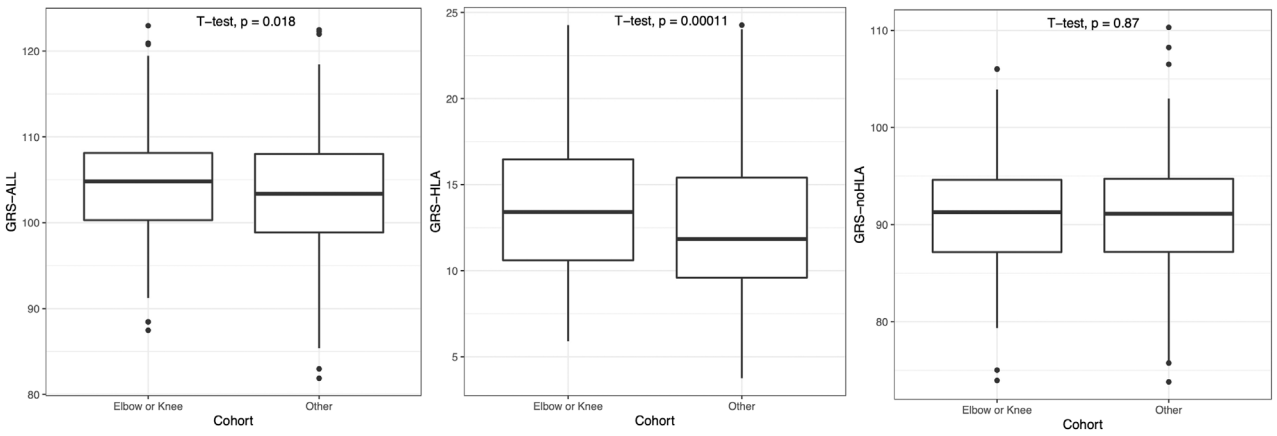


FIGURE 4
Patients presenting with psoriasis at elbow or knee at first occurrence had significantly higher GRS-ALL and GRS-HLA. Horizontal lines represent quartiles. Points represent outliers. n = 241 (Elbow or Knee), 408 (Other).

3.5 First location affected by psoriasis

A logistic regression revealed that a high GRS-ALL was significantly associated with the initial presentation of psoriasis at the elbow ($p = 0.005$) and knee ($p = 0.025$) but not with the initial presentation at the other 18 body locations (Supplementary Table S8). This association was more significant when using GRS-HLA ($p = 1.10 \times 10^{-5}$ and $p = 1.23 \times 10^{-3}$ for elbow and knee, respectively). There were no significant effects of GRS-noHLA on the first location psoriasis presented.

As psoriasis often first presents on the elbow or knee location, patients were then grouped into a combined "Elbow or Knee" *versus* "Other" category of first presentation. Patients with "Elbow or Knee" first presentation had a significantly higher GRS-ALL and GRS-HLA compared to those who had psoriasis present at other locations first ($p =$

TABLE 1 Summary of Results (- if $p > 0.05$, + if $p < 0.05$, ++ if $p < 0.01$, +++ if $p < 0.001$, ++++ if $p < 0.0001$).

		GRS-ALL (or, 95% CI)	GRS-HLA (or, 95% CI)	GRS-noHLA (OR, 95% CI)
Family Score (n = 495)	All	+ (1.02, 1.00-1.04)	-(1.02, 0.98-1.06)	-(1.02, 1.00-1.05)
Onset (n = 571)	All	+++ (0.71, 0.59-0.85)	++++ (0.46, 0.34-0.63)	-(0.90, 0.74-1.10)
	T1P	+++ (0.82, 0.72-0.92)	+++ (0.69, 0.56-0.86)	-(0.90, 0.78-1.03)
	T2P	-(1.34, 0.92-1.96)	-(0.84, 0.44-1.61)	+ (1.54, 1.02-2.33)
Severity (n = 413)	All	+ (1.02, 1.00-1.06)	++ (1.09, 1.03-1.15)	-(1.00, 0.97-1.04)
	T1P	-(1.03, 0.99-1.06)	++ (1.09, 1.03-1.16)	-(1.00, 0.96-1.04)
	T2P	-(0.97, 0.87-1.07)	-(1.01, 0.85-1.21)	-(0.95, 0.85-1.07)
Total Locations (n = 649)	All	++++ (1.16, 1.11-1.21)	++++ (1.34, 1.24-1.45)	-(1.07, 1.02-1.13)
	T1P	++++ (1.13, 1.08-1.19)	++++ (1.34, 1.23-1.47)	-(1.04, 0.98-1.10)
	T2P	-(1.08, 0.95-1.23)	-(1.13, 0.91-1.40)	-(1.05, 0.91-1.21)
Elbow or Knee (n = 649)	All	+ (0.97, 0.94-0.99)	++++ (0.92, 0.88-0.96)	-(1.00, 0.97-1.03)
Genital Psoriasis (n = 649)	All	-(0.96, 0.93-0.99)	-(0.96, 0.91-1.00)	+ (0.97, 0.94-1.00)

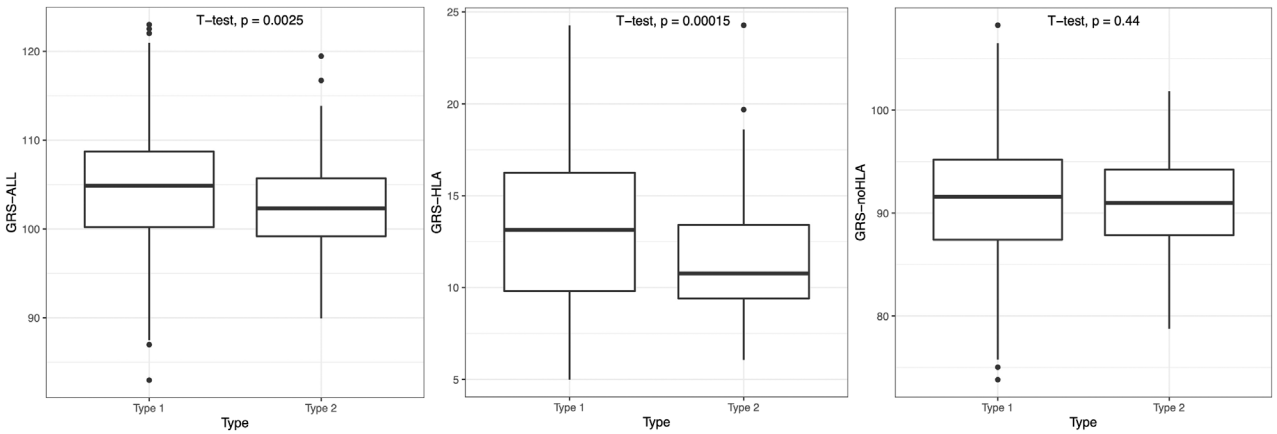


FIGURE 5
T1P patients had significantly higher GRS-ALL and GRS-HLA. Horizontal lines represent quartiles. Points represent outliers. n = 491 (T1P), 80 (T2P).

0.018 and $p = 0.00011$ respectively; Figure 4). Meanwhile, there was no significant difference in GRS-noHLA between those that presented at the "Elbow or Knee" first and those that presented at other locations first. A logistic regression was also performed to evaluate the effect of GRS on the outcome "Elbow or Knee" or "Other". GRS-ALL and GRS-HLA were significantly associated with presentation at the "Elbow or Knee" first ($p = 0.019$, OR 0.97 [0.94-0.99] and $p = 0.0001$, OR 0.92 [0.88-0.96] respectively) while GRS-noHLA was not (Table 1). Together, these results suggest that HLA SNPs are driving the first presentation of psoriasis at the elbow or knee.

3.6 Type 1 vs. type 2 psoriasis

The psoriasis subjects were split into an early-onset (age of onset <40) T1P and a late-onset (age of onset ≥ 40) T2P group, and T1P had significantly higher GRS than T2P when using GRS-ALL and GRS-HLA ($p = 0.0025$ and $p = 0.00015$ respectively; Figure 5). A linear regression analysis was performed within each onset group. In T1P, there was a significant negative association between GRS-ALL (coefficient = −0.203, $R^2 = 0.0206$, $p = 0.00084$, SE = 0.06, n = 487, OR 0.82 [0.59-0.85]). This was repeated with GRS-HLA which yielded a stronger and more significant association (coefficient = −0.367, R2 = 0.0216, $p = 0.00067$, SE = 0.11, n = 487, OR 0.69 [0.56-0.86]). When this was repeated with GRS-noHLA, the trend was weak and the association was nonsignificant (coefficient = −0.106, $R^2 = 0.0048$, $p = 0.123$, SE = 0.07, n = 487, OR 0.90 [0.78-1.03]). In T2P, there was a nonsignificant negative association between GRS-ALL and GRS-HLA with onset; however, there was a significant positive correlation between GRS-noHLA and onset (coefficient = 0.434, $R^2 = 0.041$, $p = 0.039$, SE = 0.21, n = 79). These findings suggest an association between non-HLA SNPs and age of onset in the late-onset psoriasis group.

Within the T1P and T2P onset groups, the association between GRS and severity was also analyzed. An ordinal logistic regression

was performed, and the only significant association was found in T1P using GRS-HLA (SE = 0.031, p = 0.049, OR 1.09 [1.03-1.16]). When comparing the GRS-HLA between the severity levels with a Kruskal–Wallis test followed by a pairwise Wilcoxon rank sum test, there was a significantly higher GRS in the moderate compared to the mild group within T1P (p = 0.001, Figure 3B). There was no clear trend and a nonsignificant association in T1P using GRS-ALL and GRS-noHLA an in T2P using all three GRS.

The association between GRS and total number of psoriasis locations was also analyzed within the onset groups. A linear regression analysis revealed a significant association in T1P using GRS-ALL (p = 1.40 × 10^{-6}, OR 1.13 [1.08-1.19]; Supplementary Figure S1A). This association was stronger and more significant using GRS-HLA (coefficient = 0.124, R2 = 0.045, p = 1.401 × 10^{-6}, SE = 0.025, n = 487, OR 1.34 [1.23-1.47]; Supplementary Figure S1B) while it was a nonsignificant positive association using GRS-noHLA (Supplementary Figure S1C). The trend was similar but nonsignificant in T2P for all three GRS. Overall, these results indicate that GRS is inversely related with age of onset in the T1P group while not significantly related to age of onset in the T2P group. The GRS was positively associated with psoriasis severity and total locations ever affected, specifically driven by the HLA SNPs within the early onset T1P group.

4 Discussion

In this study, we calculated the GRS-ALL, GRS-HLA, and GRS-noHLA and evaluated their correlation to psoriasis clinical features. Of note, among all psoriasis subjects, the mean GRS-ALL was 102.8, mean GRS-HLA was 12.1, and mean GRS-noHLA was 90.6. The fact that GRS-HLA is not the major quantitative contributor to GRS-ALL indicates that associations found significant for GRS-HLA but not for GRS-noHLA were not simply driven by statistical power.

Our data show that patients with a higher GRS-ALL had a higher family score; however, this association is nonsignificant when the GRS was split into GRS-HLA and GRS-noHLA suggesting that all SNPs regardless of loci cumulatively contribute to the family history. This is an interesting finding due to prior research indicating a strong association of family history with *HLA-C06:02* (Schmitt-Egenolf et al., 1993; Schmitt-Egenolf et al., 1996; Gudjonsson et al., 2002). We also found that patients with higher GRS-ALL had an earlier age of onset, and this association was found to be stronger and more significant with the GRS-HLA. This agrees with prior research showing that *HLA-C06:02* -positive patients show an earlier disease onset (Schmitt-Egenolf et al., 1993; Schmitt-Egenolf et al., 1996; Enerback et al., 1997; Gudjonsson et al., 2002). Our data also showed a significant association of GRS-ALL with psoriasis severity level, categorized as mild, moderate, and severe. This association was stronger when using GRS-HLA, suggesting that HLA SNPs could be driving the severity level. Prior studies showed an increasing severity in *HLA-Cw1*-and *HLA-Cw12*-positive patients (Onsun et al., 2019; Huang and Tsai, 2021). Interestingly, the moderate group showed a higher average GRS than the severe group which could be due to a variety of reasons. Besides random noise from sampling, this could in theory be due to specific rare psoriasis mutations, not captured by the common variants on SNP arrays, being responsible for the more severe cases of psoriasis.

We also explored the effects of GRS on the body locations ever affected and first affected by psoriasis. Our data showed that patients with a higher GRS-ALL had higher total locations affected by psoriasis. This association was more significant using GRS-HLA, suggesting that HLA SNPs could be driving this. GRS-HLA had a significant effect on the appearance of psoriasis on the hand, toe, nail, face, knee, back, chest, trunk which was unique to GRS-HLA. A prior study showed that *HLA-C*06:02*-positive patients had more extensive plaques on their arms, legs, and trunk (Gudjonsson et al., 2002). Interestingly, GRS-noHLA had a unique significant effect on the appearance of psoriasis in the genital area. While there has not yet been significant investigation into the pathology of genital or inverse psoriasis, it is known to display features atypical for plaque psoriasis (Knabel and Mudaliar, 2022). There has also been recent identification of rare gene variants in inverse psoriasis patients that have not previously been reported for psoriasis (Goblos et al., 2021).

In our follow-up analysis to explore the association between non-HLA SNPs and genital psoriasis, *IFIH1* (rs3747517) and *UBE2L3* (rs2256609) were found to contribute to genital psoriasis. Meanwhile, *IFIH1* (rs2111485) and *ERAP1* (rs39841) were found to contribute to non-genital psoriasis. Interestingly, a prior study showed that *IFIH1* is important for immune responses to *Candida* fungal infections (Jaeger et al., 2015). In addition, prior studies have associated *IFIH1* (rs3747517) and *UBE2L3* with an inability to clear hepatitis viruses, leading to chronic infection and increased inflammatory states (Liu et al., 2018; Zhu et al., 2019; Yao et al., 2021). *IFIH1* encodes a retinoic acid-inducible gene I (RIG-I)-like receptor that can sense viral RNA in order to establish a proper antiviral host response. UBE2L3 is an E2 ubiquitin conjugating enzyme that aids in regulation of many signaling pathways including NF-κB. *IFIH1* (rs2111485) and *ERAP1,* on the other hand, have previously been associated with spontaneous clearance of hepatitis virus and more efficient antigen processing during COVID-19 infection, respectively, allowing for faster resolution of inflammation (Jiang et al., 2019; D'Amico et al., 2021; Yao et al., 2021). ERAP1 is a key component of MHC class I antigen processing and presentation. Given the diverse and abundant microbiota in the genital area, we hypothesize that inflammation mediated by innate anti-fungal or anti-viral mechanisms could be correlated with the development of genital psoriasis.

Our data also showed that GRS-ALL was significantly associated with the initial presentation of psoriasis at the elbow and knee. This association became more significant when using GRS-HLA, suggesting that HLA variants could be driving this initial presentation of psoriasis at the elbow and knee. It is commonly observed that psoriasis usually first affects the elbows and knees; however, to date, there have been no studies correlating this presentation with genetic susceptibility loci partitioned between the HLA and non-HLA regions as we have done here.

Our study included an in-depth analysis of the effect of GRS within psoriasis types (early-onset T1P vs late-onset T2P). We found that the T1P group had a significantly higher GRS-ALL and GRS-HLA compared to the T2P group. We also found that T1P patients with a higher GRS-ALL and GRS-HLA had an earlier age of onset. Meanwhile, this association was nonsignificant in the T2P patients. Our data also showed that the GRS-HLA was significantly associated with severity in the T1P patients. Additionally, a higher GRS-ALL and GRS-HLA were significantly associated with more locations ever afflicted by psoriasis in the T1P patients. These findings together suggest that the GRS, driven by HLA variants, affect a

number of clinical features in T1P patients and not in T2P patients. This agrees with previous studies showing that T1P is associated with *HLA-C*06:02* while other mutations are associated with T2P (Schmitt-Egenolf et al., 1993; Hebert et al., 2015; Kepiro et al., 2021). However, extensive research on the direct effects of HLA variants on T1P subphenotypes were not reported as done here.

It is important to note that this study focused on drawing direct correlations between clinical data and HLA SNPs and non-HLA SNPs. A separate study (manuscript in preparation) discusses the power of this GRS to distinguish phenotype association across multiple cohorts. Additionally, we chose to use the most robust SNPs from a prior meta-analysis to reduce noise from less informative SNPs; however, including additional SNPs could be useful for investigating more subtle genetic associations. Other GRS tools such as PRSice2 and PRS-CS can be utilized to optimize this SNP selection. While we have presented a hypothesis on why the non-HLA SNPs that had higher contribution to genital psoriasis, further studies are needed to understand what specific pathways could be driving this. With a larger dataset, additional partitions can be made to investigate the role of specific immune pathways.

In summary, we examined the relationship between GRS-ALL, GRS-HLA, and GRS-noHLA with multiple psoriasis clinical features within a Newfoundland cohort. We found that the GRS-ALL was significantly associated with positive family history while both GRS-ALL and GRS-HLA were significantly associated with early age of onset, severity, first presentation on elbow and knee, and total locations affected, especially in the early onset T1P patients. Our findings also reveal that non-HLA SNPs influence age of onset in late-onset psoriasis as well as the presence of genital psoriasis. While some of these trends support current knowledge regarding the role of HLA in psoriasis, to our knowledge, there have not yet been studies methodically looking at non-HLA SNPs with clinical data. Nor have there been studies looking at specific body locations and their association with HLA and non-HLA SNPs. Our findings advance our knowledge of how different psoriasis susceptibility variants influence the clinical expression of disease.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://doi.org/10.6084/m9.figshare.21970847.

Ethics statement

The studies involving human participants were reviewed and approved by Newfoundland and Labrador Health Research Ethics Board (HREB) #2019.188. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

AB performed the data analysis and wrote the manuscript with support from WL, SK, JL, FO, SK, and JL performed the computations. LT assisted with methods development. WL conceived the original idea and supervised the project. All authors contributed to the article and approved the submitted version.

Acknowledgments

We are grateful to all psoriasis patients that participated in this study. We thank the late Eugene Farber for his contribution to the Newfoundland cohort.

Conflict of interest

WL has received research grant funding from Abbvie, Amgen, Janssen, Leo, Novartis, Pfizer, Regeneron, and TRex Bio. LT has received support from Galderma, Janssen, and Novartis.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1141010/full#supplementary-material

References

Abraham, G., Rutten-Jacobs, L., and Inouye, M. (2021). Risk prediction using polygenic risk scores for prevention of stroke and other cardiovascular diseases. *Stroke* 52 (9), 2983–2991. doi:10.1161/STROKEAHA.120.032619

Ahn, R., Vukcevic, D., Motyer, A., Nititham, J., Squire, D. M., Hollenbach, J. A., et al. (2021). Large-Scale imputation of KIR copy number and HLA alleles in north American and European psoriasis case-control cohorts reveals association of inhibitory KIR2DL2 with psoriasis. *Front. Immunol.* 12, 684326. doi:10.3389/fimmu.2021.684326

Alshobaili, H. A., Shahzad, M., Al-Marshood, A., Khalil, A., Settin, A., and Barrimah, I. (2010). Genetic background of psoriasis. *Int. J. Health Sci. (Qassim)* 4 (1), 23–29.

Capon, F. (2017). The genetic basis of psoriasis. *Int. J. Mol. Sci.* 18 (12), 2526. doi:10.3390/ijms18122526

Cargill, M., Schrodi, S. J., Chang, M., Garcia, V. E., Brandon, R., Callis, K. P., et al. (2007). A large-scale genetic association study confirms IL12B and leads to the identification of IL23R as psoriasis-risk genes. *Am. J. Hum. Genet.* 80 (2), 273–290. doi:10.1086/511051

Chandran, V., and Raychaudhuri, S. P. (2010). Geoepidemiology and environmental factors of psoriasis and psoriatic arthritis. *J. Autoimmun.* 34 (3), J314–J321. doi:10.1016/j.jaut.2009.12.001

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4, 7. doi:10.1186/s13742-015-0047-8

Chen, H., Hayashi, G., Lai, O. Y., Dilthey, A., Kuebler, P. J., Wong, T. V., et al. (2012). Psoriasis patients are enriched for genetic variants that protect against HIV-1 disease. *PLoS Genet.* 8 (2), e1002514. doi:10.1371/journal.pgen.1002514

Chen, H., Poon, A., Yeung, C., Helms, C., Pons, J., Bowcock, A. M., et al. (2011). A genetic risk score combining ten psoriasis risk loci improves disease prediction. *PLoS One* 6 (4), e19454. doi:10.1371/journal.pone.0019454

Chen, S. F., Dias, R., Evans, D., Salfati, E. L., Liu, S., Wineinger, N. E., et al. (2020). Genotype imputation and variability in polygenic risk score estimation. *Genome Med.* 12 (1), 100. doi:10.1186/s13073-020-00801-x

D'Amico, S., Tempora, P., Lucarini, V., Melaiu, O., Gaspari, S., Algeri, M., et al. (2021). ERAP1 and ERAP2 enzymes: A protective shield for ras against COVID-19? *Int. J. Mol. Sci.* 22 (4), 1705. doi:10.3390/ijms22041705

Enerback, C., Martinsson, T., Inerot, A., Wahlstrom, J., Enlund, F., Yhr, M., et al. (1997). Evidence that HLA-Cw6 determines early onset of psoriasis, obtained using sequence-specific primers (PCR-SSP). *Acta Derm. Venereol.* 77 (4), 273–276. doi:10.2340/0001555577273276

Goblos, A., Varga, E., Farkas, K., Arvai, K., and Kemeny, L. (2021). Genetic investigation of inverse psoriasis. *Life (Basel)* 11 (7), 654. doi:10.3390/life11070654

Goldstein, B. A., Yang, L., Salfati, E., and Assimes, T. L. (2015). Contemporary considerations for constructing a genetic risk score: An empirical approach. *Genet. Epidemiol.* 39 (6), 439–445. doi:10.1002/gepi.21912

Gudjonsson, J. E., Karason, A., Antonsdottir, A. A., Runarsdottir, E. H., Gulcher, J. R., Stefansson, K., et al. (2002). HLA-Cw6-positive and HLA-Cw6-negative patients with Psoriasis vulgaris have distinct clinical features. *J. Invest. Dermatol* 118 (2), 362–365. doi:10.1046/j.0022-202x.2001.01656.x

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., et al. (2019). The REDCap consortium: Building an international community of software platform partners. *J. Biomed. Inf.* 95, 103208. doi:10.1016/j.jbi.2019.103208

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., and Conde, J. G. (2009). Research electronic data capture (REDCap)-a metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inf.* 42 (2), 377–381. doi:10.1016/j.jbi.2008.08.010

Hebert, H. L., Bowes, J., Smith, R. L., Flynn, E., Parslew, R., Alsharqi, A., et al. (2015). Identification of loci associated with late-onset psoriasis using dense genotyping of immune-related regions. *Br. J. Dermatol* 172 (4), 933–939. doi:10.1111/bjd.13340

Henseler, T., and Christophers, E. (1985). Psoriasis of early and late onset: Characterization of two types of psoriasis vulgaris. *J. Am. Acad. Dermatol* 13 (3), 450–456. doi:10.1016/s0190-9622(85)70188-0

Huang, Y. W., and Tsai, T. F. (2021). HLA-Cw1 and psoriasis. *Am. J. Clin. Dermatol* 22 (3), 339–347. doi:10.1007/s40257-020-00585-1

Jaeger, M., van der Lee, R., Cheng, S. C., Johnson, M. D., Kumar, V., Ng, A., et al. (2015). The RIG-I-like helicase receptor MDA5 (IFIH1) is involved in the host defense against Candida infections. *Eur. J. Clin. Microbiol. Infect. Dis.* 34 (5), 963–974. doi:10.1007/s10096-014-2309-2

Jiang, W., Zhuo, L., Wu, X., Shen, J., Ye, X., Wang, Y., et al. (2019). Association of rs2111485 and rs1990760 polymorphisms of interferon induced with helicase C domain 1 gene with hepatitis C virus clearance in Chinese han population. *Viral Immunol.* 32 (8), 341–347. doi:10.1089/vim.2019.0109

Kepiro, L., Szell, M., Kovacs, L., Keszthelyi, P., Kemeny, L., and Gyulai, R. (2021). The association of HLA-C and ERAP1 polymorphisms in early and late onset psoriasis and psoriatic arthritis patients of Hungary. *Postepy Dermatol Alergol.* 38 (2), 43–51. doi:10.5114/ada.2021.104277

Knabel, M., and Mudaliar, K. (2022). Histopathologic features of inverse psoriasis. *J. Cutan. Pathol.* 49 (3), 246–251. doi:10.1111/cup.14142

Liu, Y., Song, C., Ni, H., Jiao, W., Gan, W., Dong, X., et al. (2018). UBE2L3, a susceptibility gene that plays oncogenic role in hepatitis B-related hepatocellular carcinoma. *J. Viral Hepat.* 25 (11), 1363–1371. doi:10.1111/jvh.12963

Lu, Y., Kane, S., Chen, H., Leon, A., Levin, E., Nguyen, T., et al. (2013). The role of 39 psoriasis risk variants on age of psoriasis onset. *ISRN Dermatol* 2013, 203941. doi:10.1155/2013/203941

Machiela, M. J., and Chanock, S. J. (2015). LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 31 (21), 3555–3557. doi:10.1093/bioinformatics/btv402

Meigs, J. B., Shrader, P., Sullivan, L. M., McAteer, J. B., Fox, C. S., Dupuis, J., et al. (2008). Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N. Engl. J. Med.* 359 (21), 2208–2219. doi:10.1056/NEJMoa0804742

Nall, L., Gulliver, W., Charmley, P., and Farber, E. M. (1999). Search for the psoriasis susceptibility gene: The Newfoundland study. *Cutis* 64 (5), 323–329.

Onsun, N., Pirmit, S., Ozkaya, D., Celik, S., Rezvani, A., Cengiz, F. P., et al. (2019). The HLA-cw12 allele is an important susceptibility allele for psoriasis and is associated with resistant psoriasis in the Turkish population. *ScientificWorldJournal* 2019, 7848314. doi:10.1155/2019/7848314

Papp, K., Gulliver, W., Lynde, C., Poulin, Y., Ashkenas, J., and Canadian Psoriasis Guidelines, C. (2011). Canadian guidelines for the management of plaque psoriasis: Overview. *J. Cutan. Med. Surg.* 15 (4), 210–219. doi:10.2310/7750.2011.10066

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). Plink: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81 (3), 559–575. doi:10.1086/519795

Rahman, P., Jones, A., Curtis, J., Bartlett, S., Peddle, L., Fernandez, B. A., et al. (2003). The Newfoundland population: A unique resource for genetic investigation of complex diseases. *Hum. Mol. Genet.* 2, R167–R172. doi:10.1093/hmg/ddg257

Schmitt-Egenolf, M., Boehncke, W. H., Stander, M., Eiermann, T. H., and Sterry, W. (1993). Oligonucleotide typing reveals association of type I psoriasis with the HLA-DRB1*0701/2, -DQA1*0201, -DQB1*0303 extended haplotype. *J. Invest. Dermatol* 100 (6), 749–752. doi:10.1111/1523-1747.ep12476080

Schmitt-Egenolf, M., Eiermann, T. H., Boehncke, W. H., Stander, M., and Sterry, W. (1996). Familial juvenile onset psoriasis is associated with the human leukocyte antigen (HLA) class I side of the extended haplotype Cw6-B57-drb1*0701-dqa1*0201-dqb1*0303: A population- and family-based study. *J. Invest. Dermatol* 106 (4), 711–714. doi:10.1111/1523-1747.ep12345600

Stuart, P. E., Nair, R. P., Tsoi, L. C., Tejasvi, T., Das, S., Kang, H. M., et al. (2015). Genome-wide association analysis of psoriatic arthritis and cutaneous psoriasis reveals differences in their genetic architecture. *Am. J. Hum. Genet.* 97 (6), 816–836. doi:10.1016/j.ajhg.2015.10.019

Tsoi, L. C., Stuart, P. E., Tian, C., Gudjonsson, J. E., Das, S., Zawistowski, M., et al. (2017). Large scale meta-analysis characterizes genetic architecture for common psoriasis associated variants. *Nat. Commun.* 8, 15382. doi:10.1038/ncomms15382

Weedon, M. N., McCarthy, M. I., Hitman, G., Walker, M., Groves, C. J., Zeggini, E., et al. (2006). Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. *PLoS Med.* 3 (10), e374. doi:10.1371/journal.pmed.0030374

Wellcome Trust Case Control, C. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447 (7145), 661–678. doi:10.1038/nature05911

Yanovsky, R. L., Chen, H., Leslie, S., Carrington, M., and Liao, W. (2020). The interaction of LILRB2 with HLA-B is associated with psoriasis susceptibility. *J. Invest. Dermatol* 140 (6), 1292–1295. doi:10.1016/j.jid.2019.12.006

Yao, Y., Shen, Y., Shao, H., Liu, Y., Ji, Y., Du, G., et al. (2021). Polymorphisms of RIG-I-like receptor influence HBV clearance in Chinese Han population. *J. Med. Virol.* 93 (8), 4957–4965. doi:10.1002/jmv.26969

Zeng, X., Chen, H., Gupta, R., Paz-Altschul, O., Bowcock, A. M., and Liao, W. (2013). Deletion of the activating NKG2C receptor and a functional polymorphism in its ligand HLA-E in psoriasis susceptibility. *Exp. Dermatol* 22 (10), 679–681. doi:10.1111/exd.12233

Zhu, X. B., Zhuo, L. Y., Yue, M., Liu, M., Zang, F., Fan, H. Z., et al. (2019). Genetic variants in IFIH1 and DDX58 influence hepatitis C virus clearance in Chinese Han population. *J. Med. Virol.* 91 (6), 1097–1103. doi:10.1002/jmv.25398

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 21 June 2023
DOI 10.3389/fgene.2023.1173676

Check for updates

OPEN ACCESS

EDITED BY
Xianyong Yin,
University of Michigan, Ann Arbor,
United States

REVIEWED BY
Ramcés Falfán-Valencia,
National Institute of Respiratory
Diseases-Mexico (INER), Mexico
Andrew Dewan,
Yale University, United States

*CORRESPONDENCE
Alison A. Motsinger-Reif,
alison.motsinger-reif@nih.gov

RECEIVED 24 February 2023
ACCEPTED 09 June 2023
PUBLISHED 21 June 2023

CITATION
Lee EY, Choi W, Burkholder AB, Perera L, Mack JA, Miller FW, Fessler MB, Cook DN, Karmaus PWF, Nakano H, Garantziotis S, Madenspacher JH, House JS, Akhtari FS, Schmitt CS, Fargo DC, Hall JE and Motsinger-Reif AA (2023), Race/ethnicity-stratified fine-mapping of the MHC locus reveals genetic variants associated with late-onset asthma. *Front. Genet.* 14:1173676. doi: 10.3389/fgene.2023.1173676

COPYRIGHT
© 2023 Lee, Choi, Burkholder, Perera, Mack, Miller, Fessler, Cook, Karmaus, Nakano, Garantziotis, Madenspacher, House, Akhtari, Schmitt, Fargo, Hall and Motsinger-Reif. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Race/ethnicity-stratified fine-mapping of the MHC locus reveals genetic variants associated with late-onset asthma

Eunice Y. Lee[1], Wonson Choi[2], Adam B. Burkholder[3], Lalith Perera[4], Jasmine A. Mack[1,5], Frederick W. Miller[6], Michael B. Fessler[7], Donald N. Cook[7,8], Peer W. F. Karmaus[7], Hideki Nakano[7], Stavros Garantziotis[9], Jennifer H. Madenspacher[9], John S. House[1], Farida S. Akhtari[1,9], Charles S. Schmitt[10], David C. Fargo[3], Janet E. Hall[9] and Alison A. Motsinger-Reif[1]*

[1]Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, Durham, NC, United States, [2]Genomics and Bioinformatics Laboratory, Seoul National University, Seoul, Republic of Korea, [3]National Institute of Environmental Health Sciences, Durham, NC, United States, [4]Genomic Integrity and Structural Biology Laboratory, National Institute of Environmental Health Sciences, Durham, NC, United States, [5]Department of Obstetrics and Gynecology, University of Cambridge, Cambridge, United Kingdom, [6]Environmental Autoimmunity Group, Clinical Research Branch, National Institute of Environmental Health Sciences, Durham, NC, United States, [7]Immunity, Inflammation and Disease Laboratory, National Institute of Environmental Health Sciences, Durham, NC, United States, [8]Immunogenetics Group, National Institute of Environmental Health Sciences, Durham, NC, United States, [9]Clinical Research Branch, National Institute of Environmental Health Sciences, Durham, NC, United States, [10]Division of Translational Toxicology, National Institute of Environmental Health Sciences, Durham, NC, United States

Introduction: Asthma is a chronic disease of the airways that impairs normal breathing. The etiology of asthma is complex and involves multiple factors, including the environment and genetics, especially the distinct genetic architecture associated with ancestry. Compared to early-onset asthma, little is known about genetic predisposition to late-onset asthma. We investigated the race/ethnicity-specific relationship among genetic variants within the major histocompatibility complex (MHC) region and late-onset asthma in a North Carolina-based multiracial cohort of adults.

Methods: We stratified all analyses by self-reported race (i.e., White and Black) and adjusted all regression models for age, sex, and ancestry. We conducted association tests within the MHC region and performed fine-mapping analyses conditioned on the race/ethnicity-specific lead variant using whole-genome sequencing (WGS) data. We applied computational methods to infer human leukocyte antigen (HLA) alleles and residues at amino acid positions. We replicated findings in the UK Biobank.

Results: The lead signals, rs9265901 on the 5′ end of HLA-B, rs55888430 on HLA-DOB, and rs117953947 on HCG17, were significantly associated with late-onset asthma in all, White, and Black participants, respectively (OR = 1.73, 95%CI: 1.31 to 2.14, $p = 3.62 \times 10^{-5}$; OR = 3.05, 95%CI: 1.86 to 4.98, $p = 8.85 \times 10^{-6}$; OR = 19.5, 95%CI: 4.37 to 87.2, $p = 9.97 \times 10^{-5}$, respectively). For the HLA analysis, HLA-B*40:02 and HLA-DRB1*04:05, HLA-B*40:02, HLA-C*04:01, and HLA-DRB1*04:05,

and HLA-DRB1*03:01 and HLA-DQB1 were significantly associated with late-onset asthma in all, White, and Black participants.

Conclusion: Multiple genetic variants within the MHC region were significantly associated with late-onset asthma, and the associations were significantly different by race/ethnicity group.

KEYWORDS

MHC, HLA allele, immune function, race, ethnicity, late-onset asthma

1 Introduction

Asthma is a chronic disease of the airways characterized by inflammation, mucus production, and reversible airway obstruction that impairs normal breathing. According to the Centers for Disease Control and American Thoracic Society, 21 million adults in the United States have asthma, and $80 billion is spent annually on healthcare related to the disease (American Thoracic Society, 2018; Centers for Disease Control, 2022). The complex etiology of asthma involves multiple factors, including genetic and environmental risk factors.

Recent studies have revealed that asthma is not a single disease but rather comprises multiple phenotypes (Anderson, 2008; Lötvall et al., 2011). This phenotypic heterogeneity likely stems from the involvement of distinct biological mechanisms, a concept captured by the term "asthma endotypes." An improved understanding of these endotypes may reveal novel pathways that can be selectively targeted for specific forms of asthma. Our understanding of early-onset, or childhood, asthma has improved, and genetic predisposition is known to play an important role. In general, individuals with early-onset asthma respond well to inhaled corticosteroids, which is the gold standard for asthma treatment (Hirano and Matsunaga, 2018). By contrast, individuals with late-onset asthma are often steroid-resistant, and the potential mechanisms that lead to late-onset asthma remain poorly understood (Hirano and Matsunaga, 2018). In particular, little is known about genetic predisposition to late-onset asthma. Improved knowledge of genes that affect late-onset asthma may reveal novel biologic pathways for targeted therapies.

Genome-wide association studies (GWAS) are a fruitful approach to identifying genes associated with a wide array of diseases, including asthma. GWAS have consistently identified significant associations between various asthma phenotypes and genes within the major histocompatibility complex (MHC) and human leukocyte antigen (HLA) loci (Li et al., 2010; Galanter et al., 2014; Clay et al., 2022). The human MHC region, located on chromosome 6 with an approximate size of 150–180 Mb, is the most gene-dense area of the genome and harbors highly polymorphic genes, including HLA genes. The HLA system encodes a peptide-binding groove of the HLA molecule critical for antigen binding and T cell recognition and the subsequent activation of immune response upon antigen challenge and thus plays an essential role in adaptive immunity. However, the pathological roles of genes within the MHC region and HLA alleles in late-onset asthma remain unclear.

Prior GWAS have sometimes failed to replicate genetic associations in asthma, which has hampered efforts to delineate plausible mechanisms. This failure may be due to population heterogeneity, diverse asthma phenotypes and endotypes, and limitations of single nucleotide polymorphism (SNP)-level studies. In addition, inference of accurate high-resolution HLA alleles and amino acid residues remains challenging due to the tight linkage disequilibrium (LD) across disease-associated MHC haplotypes and the highly polymorphic nature of associated variants. To address these issues and identify putative causal variants, we conducted race/ethnicity-stratified fine-mapping studies of the MHC region for late-onset asthma utilizing WGS samples collected from a North Carolina-based multiethnic adult cohort. We also applied appropriate and high-accuracy computational methods/tools to infer HLA alleles and amino acid positions and residues for individuals of African and European ancestry using WGS data. The results show heterogeneity at three levels of resolution, namely, single nucleotide variants (SNVs), HLA alleles, and amino acids.

2 Materials and methods

2.1 Study participants: PEGS cohort

The ongoing North Carolina-based Personalized Environment and Genes Study (PEGS) (N = 19,672) began recruiting participants in 2002 and collects questionnaire-based exposome and health history data. The race/ethnicity of PEGS participants is categorized into American Indian/Alaska Native, Asian, Black, Hispanic, Native Hawaiian/Pacific Islander, White, and other/multiple races/ethnicities. However, due to small sample sizes, we focused our analyses on three racial/ethnic groups: non-Hispanic Black individuals, non-Hispanic White individuals, and all individuals. PEGS is described in detail elsewhere (Lee et al., 2022). Local institutional review boards approved the studies, and all participants provided written informed consent.

2.2 Asthma case and control definition

For the late-onset asthma outcome, we assigned participants to the case group if they answered "YES" when asked if they have ever been diagnosed with asthma (as defined by Global Initiative for Asthma criteria) (Global Initiative for Asthma, 2022) by a physician, currently have asthma, do not have other respiratory diseases, including idiopathic pulmonary fibrosis (IPF), chronic obstructive pulmonary disease (COPD), and tuberculosis (TB), and were diagnosed with asthma at the age of 20 years or more. We assigned participants to the control group if they answered "NO"

when asked if they have ever been diagnosed with asthma, have other respiratory diseases, including IPF, COPD, and TB, and have a regular cough and/or breathlessness, regular wheezing, or whistling in the chest.

2.3 MHC region-wide allelic association tests

We performed race/ethnicity-stratified MHC region-wide association testing for late-onset asthma in 3,641 PEGS participants, adjusting for age, sex, and four ancestry groups. The results of the MHC region-wide association tests for each ancestry are described below and summarized in the Supplementary Material.

2.4 Fine-mapping: conditional analysis

We built a multi-SNP logistic regression model, applying a forward selection procedure for all, White, and Black participants and adjusted for age, sex, and four ancestry groups, conditional on the lead SNP (Lee et al., 2020). Details of the fine-mapping conditional analysis can be found in the Supplementary Material (Supplementary Section S1.5; Supplementary Table S1).

2.5 Validation

To validate our results from the association tests, we examined associations of the lead SNPs for each analysis (all, White, and Black participants) in GWAS results from the Pan-UK Biobank (Pan-UK Biobank, 2022) for asthma and asthma-related phenotypes. The UK Biobank (RRID:SCR_012815) is a large-scale biomedical database with both genetic and phenotype data for several common diseases and traits for approximately 500,000 participants from the United Kingdom (Sudlow et al., 2015). The Pan-UK Biobank study conducted a multi-ancestry GWAS of 7,228 phenotypes, including asthma and asthma-related phenotypes such as allergic rhinitis, lung function, and respiratory infections. Summary statistics are available for each ancestry group and a meta-analysis across all populations.

2.6 Functional annotation of genetic variants

We conducted annotation for the lead SNP and candidate causal SNPs identified from the conditional analyses using publicly available data (Boyle et al., 2012; Carithers et al., 2015; Martin et al., 2017; Carvalho-Silva et al., 2019; Luo et al., 2020). The Supplementary Material provides more information (Supplementary Section S1.6).

2.7 Assembly of HLA alleles: Kourami

We used Kourami (RRID:SCR_022280), a graph-guided assembly technique, to generate four- and six-digit HLA alleles for six common HLA genes types (HLA-A, -B, -C, -DQA, DQB, and -DRB) and 14 additional HLA genes (HLA-DOA, -DOB, -DMA, -DMB, -DPA1, -DPB1, -DRQ, -DRB3, -DRB5, -F, -G, -H, -J, and -L) (Lee and Kingsford, 2018). The Supplementary Material provides more information on Kourami (Supplementary Section S1.7).

2.8 Amino acid position inference

In addition to HLA alleles, we inferred residues at amino acid positions. We applied the CookHLA and HLA analysis toolkit (HATK) enrichment-free computational HLA imputation/inference methods (Choi et al., 2021; Cook et al., 2021). The Supplementary Material provides details (Supplementary Section S1.8).

2.9 Structure implications

We selected the highest-resolution X-ray crystal structures for HLA-DRB (pdb ID: 4x5w, resolution 1.34A), HLA-B (pdb ID: 1K5N, resolution 1.09 A), and HLA-C (pdb ID: 6JTO, resolution 1.70A). Structures were rendered using Chimera (RRID:SCR_002959).

We conducted all statistical analyses in R Project for Statistical Computing (RRID:SCR_001905, version 4.2.1) and PLINK 2.0 (RRID:SCR_001757) (Purcell et al., 2007; Habu et al., 2014).

3 Results

3.1 Study participants

On average, asthma cases were older than controls (54.3 ± 11.1 and 50.3 ± 14.8 years, respectively: $p = 4.1 \times 10^{-4}$) (Supplementary Table S2). The proportion of males in the case group was lower than in the control group (15.5% vs. 32.8%: $p = 2.8 \times 10^{-6}$). The small differences in genomic ancestry proportions between cases and controls were not statistically significant except for Black participants (18.5% ± 32.8% and 12.6% ± 28.4%, respectively: $p = 7.3 \times 10^{-3}$). Figure 1 displays the variability of genomic ancestry proportions for individual participants, and Supplementary Figure S2 outlines the population structure of PEGS participants by self-reported race and ethnicity.

3.2 MHC region-wide association studies

We performed race/ethnicity-stratified MHC region-wide association testing for late-onset asthma in 3,641 PEGS participants, adjusting for covariates. We discovered two MHC region-wide peaks for all and White participants and a suggestive peak for Black participants after multiple testing correction (Figure 2). The most significant association for all participants is located near *HLA-B*. The lead signal, rs9265901, is on the 5' end of *HLA-B* (Figure 3). A one-unit increase in the affected allele (minor allele: G) was associated with 67% increased odds of being a case

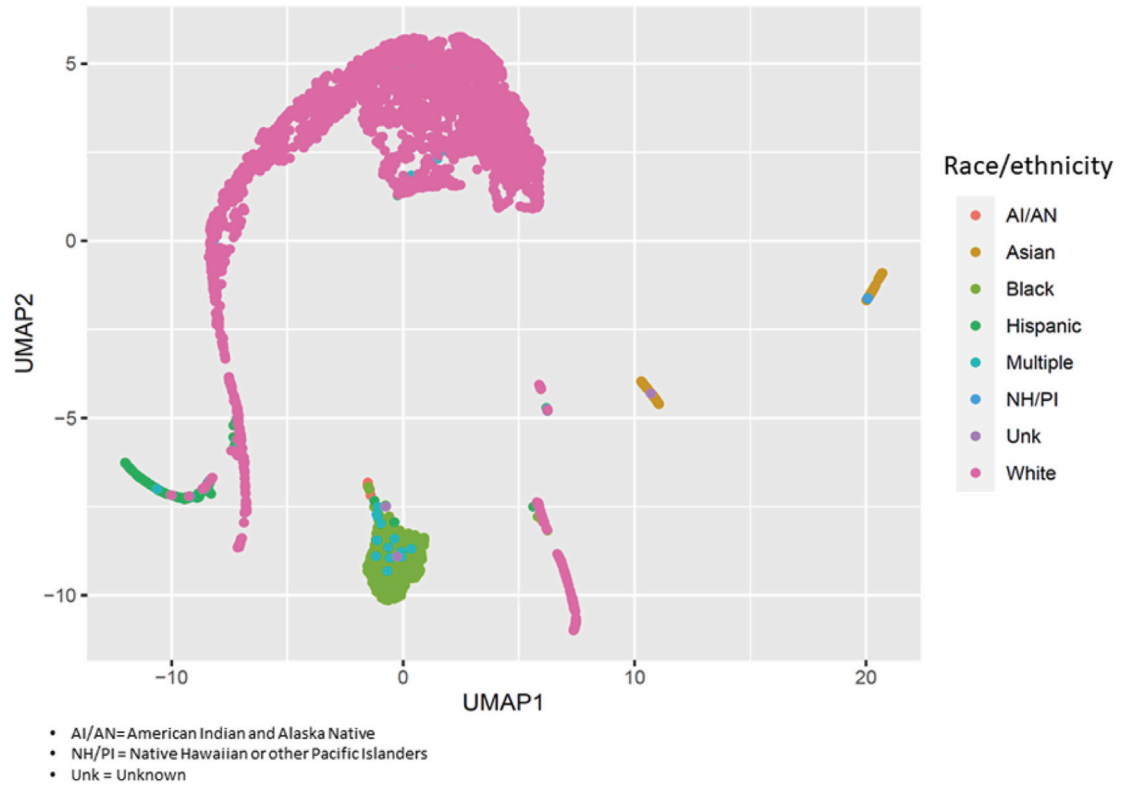


FIGURE 1
Population structure of PEGS participants. We applied uniform manifold approximation and projection (UMAP) to visualize the genetic variability of study participants using the first 40 principal components of whole-genome sequencing data. Each circle represents a study participant, and the colors represent their self-reported race/ethnicity.

(OR = 1.73, 95%CI: 1.31 to 2.14, $p = 3.62 \times 10^{-5}$). For White participants, the lead SNP, rs55888430, is on *HLA-DOB* (OR = 3.05, 95%CI: 1.86 to 4.98, $p = 8.85 \times 10^{-6}$). For Black participants, the lead SNP, rs117953947, is on *HCG17* (OR = 19.5, 95%CI: 4.37 to 87.2, $p = 9.97 \times 10^{-5}$). The lead SNPs for each reference ancestry are weakly to moderately correlated with cis-eQTLs for multiple genes in multiple tissues according to the GTEx eQTL Browser (RRID:SCR_001618) (Table 1; Figure 4) (Carithers et al., 2015). The results of the MHC region-wide association tests for each ancestry are summarized in the Supplementary Material.

3.3 Fine-mapping: conditional analysis

To identify candidate causal variants, we conducted conditional analyses to determine associations between late-onset asthma and the affected/minor allele at an SNP. Because the different allelic association signals observed for different ancestries were not due to a single SNP, we further investigated the joint effects of multiple SNPs on late-onset asthma.

Using multi-SNP logistic regression models with a forward selection procedure, we identified 10, 9, and 12 significant, independent SNPs for all, White, and Black participants, respectively (Table 2). The effect estimates in Table 2 represent the association between an SNP and late-onset asthma, conditioned on the lead SNP and adjusted for covariates and additional SNPs. The table contains only SNPs that remained significant after adjusting for additional SNPs and covariates. Table 3 displays the associations and minor allele frequencies (MAF) of the lead SNPs across all strata. The Supplementary Material provides a full list of the selected SNPs (Supplementary Table S3).

After identifying candidate causal variants from the conditional analyses, we performed functional annotation of these SNPs to elucidate their potential biological roles in late-onset asthma in PEGS participants. Supplementary Table S4 is a full list of candidate causal variants. Based on annotation from publicly available data, three SNPs identified for all participants exhibit a regulatory function in white blood cells and lung tissue (Supplementary Table S5). rs4569, a 3′UTR variant of *LY6G5B*, was strongly correlated with the negative expression of *LY6G5C* in whole blood and lung tissue ($p = 1.3 \times 10^{-119}$ and $p = 1.8 \times 10^{-43}$, respectively) (Supplementary Table S4). A 5′UTR variant of *BRD2*, rs62407970, intersected with strong promoter-like signatures, including a DNase I hypersensitive site and a histone modification (H3K4me3 and H3K27ac ChIP-Seq) (cCRE accession ID: EH38E3701714). Multiple experimental data indicate that rs62407970 falls within the *POLR2A* transcription factor binding

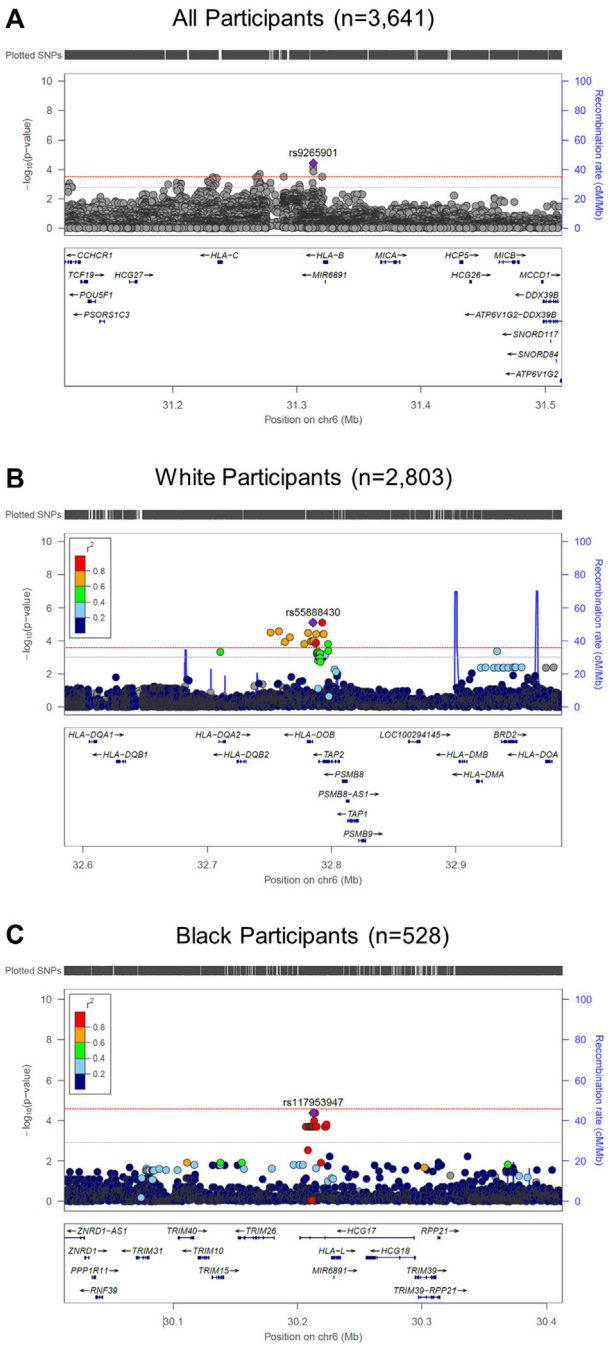


FIGURE 2
LocusZoom (RRID:SCR_009257) plots of MHC region-wide studies. **(A)** For all participants (n = 3,641), rs9265901 (overlapping gene: HLA-B) was associated with late-onset asthma; **(B)** For White participants, rs55888430, an intron variant of HLA-DOB, was significantly associated with late-onset asthma; **(C)** For Black participants, rs117953947, which overlaps with HCG17, was significantly associated with late-onset asthma. We adjusted all regression models for age, sex, and four ancestries. The red and gray dotted lines represent the significance and suggestive thresholds.

site and may interact with *HLA-DPB2* in human lung fibroblast cells (IMR90) and mesenchymal stem cells (MES) (Boyle et al., 2012).

For White participants, the non-coding exon variant of *HCG24*, rs7754362, intersected with low DNase-seq and H3K27ac ChIP-Seq peaks in lung tissue and CD14$^+$ monocytes and high CTCF signals in human lung fibroblast cells (cCRE accession ID: EH38E3701814). In addition, ChIP-Seq experiments indicate that rs7754362 is located within a CTCF transcription factor binding site in blood and lung tissue. Long-range chromatic interaction data suggest that rs7754362 may interact with *HLA-DMA* and *BRD2* in IMR90 and MES cell lines. Within the genomic region identified for Black participants, the intron variant rs17198965 overlapped with strong promoter-like signatures, including high DNase-seq and H3K4me3 ChIP-Seq peaks. Experimental data suggest that rs17198965 resides in the binding site of transcription factor *EZH2* (cCRE accession ID: EH38E3700474).

3.4 Validation of lead SNPs

We examined the associations of the lead SNPs from our association analyses with asthma and asthma-related phenotypes in the UK Biobank data using Pan-UK Biobank GWAS results (Pan-UK Biobank, 2022). We found that the lead SNP for all participants, *rs9265901*, is associated with asthma (OR = 1.04, $p = 3.36 \times 10^{-4}$), atopic dermatitis (OR = 1.12, $p = 4.33 \times 10^{-5}$), decreased lung function (βeta = −0.01, $p = 1.47 \times 10^{-5}$), and allergic rhinitis (OR = 1.04, $p = 5.62 \times 10^{-6}$) in the UK Biobank (Raj et al., 2014). In addition, *rs9265901* is associated with increased lymphocytes (βeta = 0.06, $p = 2.35 \times 10^{-135}$), eosinophils (βeta = 0.03, $p = 1.68 \times 10^{-25}$), and neutrophils (βeta = 0.06, $p = 1.33 \times 10^{-103}$) (Raj et al., 2014). The top SNP for White participants, *rs55888430*, is associated with increased odds of being an asthma case (OR = 1.14, $p = 1.78 \times 10^{-4}$), having upper respiratory infections (OR = 1.30, $p = 1.95 \times 10^{-3}$), and having increased eosinophils (βeta = 0.05, $p = 9.82 \times 10^{-16}$) (Poplin et al., 2018). However, the association of this variant with reduced odds of both late- and early-onset asthma has been reported elsewhere (OR = 0.39, $p = 3.2 \times 10^{-3}$; OR = 0.41, $p = 2.0 \times 10^{-4}$, respectively) (Poplin et al., 2018). Lastly, the lead variant for Black participants, *rs117953647*, is associated with opportunistic respiratory infection (OR = 1.71, $p = 1.82 \times 10^{-4}$), pneumonia (OR = 2.55, $p = 2.02 \times 10^{-3}$), and increased lymphocytes (βeta = 0.03, $p = 2.15 \times 10^{-4}$) (Manichaikul et al., 2010), further validating our findings and providing evidence supportive of the proposed ancestry-specific pathways.

3.5 HLA analysis: HLA alleles

For all participants, several classical HLA alleles were significantly associated with late-onset asthma after adjusting for covariates (Table 4). The strongest association was *HLA-B*40:02* (OR = 3.18, 95%CI: 1.63 to 6.18), which is concordant with the MHC-wide association test result indicating the lead SNP (rs9265901) is located near *HLA-B* (Supplementary Figure S3). The second strongest association was *HLA-DRB1*04:05* (OR = 6.94, 95%CI: 2.15 to 22.44). For White participants, two MHC class I alleles, *HLA-B*40:02* and *HLA-C*04:01*, and one MHC class II allele, *HLA-DRB1*04:05*, were significantly associated with late-onset asthma. *HLA-C*04:01* was protective for being a late-onset asthma case (OR = 0.25, 95%CI: 0.09 to 0.70) whereas *HLA-B*40:02* and *HLA-DRB1*04:05* increased the odds of being a case [(OR = 3.77, 95%CI: 1.91 to 7.44; OR = 8.19, 95%CI: 2.51 to 26.79), respectively]. For Black participants, two MHC class II

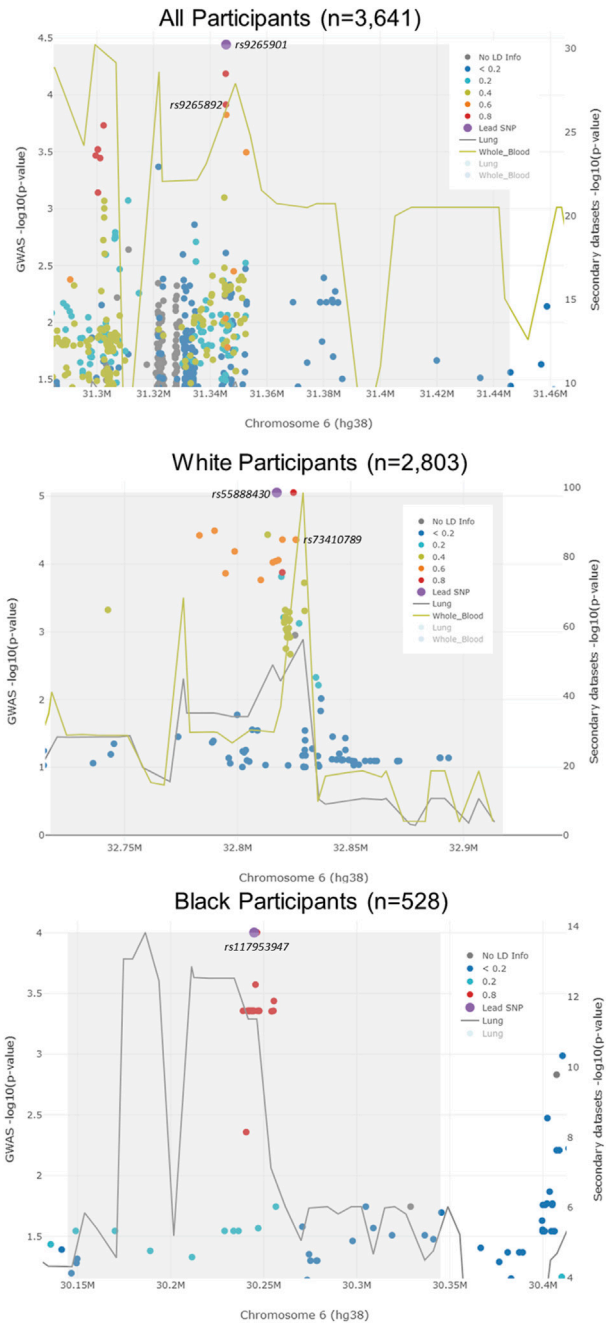


FIGURE 3
LocusFocus plots of colocalization analysis results. The lead SNP for all participants (rs9265901) was not associated with the expression of HLA-B in lung tissue or whole blood. However, an SNP that was highly correlated with the lead SNP (rs9265892) was associated with the expression of HLA-B in whole blood. The lead SNP (rs55888430) for White participants was not associated with the expression of HLA-DOB. An LD SNP for the lead SNP (rs73410789) was associated with HLA-DOB expression in whole blood. The lead SNP for Black participants, rs117953947, was not associated with the expression of HCG17 in lung tissue.

alleles, *HLA-DRB1*03:01* and *HLA-DQB1*02:01*, were significantly associated with increased risk of being a late-onset asthma case [(OR = 4.00, 95%CI: 1.46 to 10.91; OR = 2.25, 95%CI: 1.09 to 4.65), respectively].

3.6 HLA analysis

In accordance with the HLA allele analysis results, within *HLA-DRB1*, the amino acid variants at positions 37 and 38 were significantly associated with late-onset asthma. Among the three allelic variants at position 38 (leucine, valine, alanine) and five variants at position 37 (serine, phenylalanine, tyrosine, asparagine, leucine), leucine was the strongest risk residue at both positions (Supplementary Figure S4E). Within *HLA-C*, threonine residue at position 73 and alanine residue at position 90 were associated with increased odds of being an asthma case (OR = 1.40, 95%CI: 1.13 to 1.74; OR = 1.44, 95%CI: 1.16 to 1.79, respectively). For White participants, the leucine allele at amino acid 103 within *HLA-B* was associated with reduced odds of being a late-onset asthma case (OR = 0.48, 95%CI: 0.30 to 0.74). For Black participants, threonine residue at position 73 within *HLA-C* was associated with a 126% increase in the odds of being an asthma case.

The crystal structures in Figure 5 demonstrate that *HLA-DRB1* residues at 37 and 38 are located on the peptide-binding surface. Residue 103 within *HLA-B*, which appears to interact with a peptide, may play an important role in maintaining the structural scaffold and creating a binding surface. Within *HLA-C*, residues 73 and 90 are located on the peptide-binding groove.

4 Discussion

4.1 Findings

Previous epidemiological studies and GWAS have identified potential biological mechanisms underlying asthma, but few studies have focused on late-onset asthma. Prior GWAS have identified variants within the MHC region that are significantly associated with various asthma phenotypes (Li et al., 2010; Galanter et al., 2014). The MHC locus is the most gene-diverse and gene-dense region of the human genome, and the genetic architecture of the region is shaped by multiple factors, including gene shuffling and selective pressure in response to migration-related exposure to various environmental pathogens (Prugnolle et al., 2005; Traherne et al., 2006). We propose that racial and ethnic differences in allele frequencies in the MHC region of the genome may contribute, at least in part, to the observed racial disparities in asthma.

We leveraged a multi-ancestry cohort and WGS data to identify race/ethnicity-specific genetic variants associated with late-onset asthma. This is one of only a few studies that have conducted race/ethnicity-specific fine-mapping of the MHC region in the context of late-onset asthma in a United States-based cohort (Daya et al., 2021). We discovered several candidate gene variants associated with late-onset asthma that appear to play roles in allergy- and atopy-mediated airway inflammation and hyperresponsiveness. Additionally, we discovered race/ethnicity-specific candidate variants involved in immune pathways.

As additional validation of our findings, the lead variants we discovered are also well-established expressed quantitative trait loci (eQTL). For all participants, we discovered variants associated with both innate and adaptive immunity, primarily allergy- and atopy (eczema)-prone pathways (Figure 6). The 3′UTR *LY6G5C* variant *rs4569* is inversely associated with expression of *LY6G5C* and

TABLE 1 Amino acids analysis results for all, White, and Black participants. HLA amino acid positions with a particular residue were associated with late-onset asthma. We adjusted all regression models for age, sex, and four ancestries.

HLA amino acids analysis						
All participants						
HLA-(gene)	AA position	Genetic position	Residue	OR (95% CI)		*p*-value
C	90	31347357_exon2	A/D	1.40 (1.13 to 1.74)		2.34×10^{-3} ǂ
C	73	31347408_exon2	T/A	1.44 (1.16 to 1.79)		9.90×10^{-4} ǂ
DRB1	38	32660026_exon2	L	2.59 (1.56 to 4.28)		2.16×10^{-4} ǂ
DRB1	38	32660026_exon2	V	2.04 (1.26 to 3.30)		3.64×10^{-3} ǂ
DRB1	37	32660029_exon2	L	2.59 (1.56 to 4.28)		2.16×10^{-4} ǂ
White participants						
B	103	31432162_exon3	L/V	0.48 (0.30 to 0.74)		1.08×10^{-3} §
Black participants						
C	73	31347408_exon2	T/A	2.26 (1.38 to 3.72)		1.31×10^{-3} ƒ

Amino acid position analysis: ǂPooled ancestry: significant *p*-value = 3.9×10^{-4}; suggestive *p*-value = 3.9×10^{-3}. §EUR ancestry: significant *p*-value = 5.6×10^{-4}; suggestive *p*-value = 5.6×10^{-3}. ƒAFR ancestry: significant *p*-value = 3.4×10^{-4}; suggestive *p*-value = 3.4×10^{-3}.

LY6G5B, which are members of the lymphocyte antigen superfamily *LY6/uPAR* (Loughner et al., 2016), both in lung tissue ($p = 1.8 \times 10^{-43}$; $p = 1.2 \times 10^{-29}$, respectively) and whole blood ($p = 1.3 \times 10^{-119}$; $p = 1.8 \times 10^{-33}$, respectively) (Supplementary Table S4). *LY6* protein has previously been characterized as playing a critical role in inflammatory cell regulation, including activation, proliferation, migration, interactions between cells, maturation of antigen-presenting cells (dendritic cells and macrophages), and induction of cytokines upon antigen challenges due to infection and various environmental stimuli (Lee et al., 2013; Loughner et al., 2016). Additionally, publicly available chromatin conformation data (Hi-C) suggest that *rs4569* is in a long-range distal region that interacts with the *NEU1* gene in mesenchymal stem cells (MES) (Supplementary Table S5). The enzymatic activity of *NEU1* increases the signaling of T helper 2 (T_H2) cells and was found to infiltrate and accumulate in the airways and inflamed areas in murine models during acute asthma attacks and children with asthma after viral infection (Katoh et al., 2010; Pech et al., 2018).

For White participants, SNPs and HLA alleles identified from fine-mapping of the MHC region were also associated with genes that regulate innate and adaptive immune responses, particularly pathways involving atopy-related and allergen-driven airway inflammation and hyperresponsiveness (Figure 6). The lead SNP, *rs55888430*, has been shown to interact with the butyrophilin-like 2 (*BTNL2*) gene in IMR90 and MES cell lines (Supplementary Table S5) (Martin et al., 2017). *BTNL2* has also been associated with atopic asthma in children in Korea, dust-mite-specific IgE response in a Japanese population, and sarcoidosis in a White German population (Rybicki et al., 2005; Konno et al., 2009; Kim et al., 2021). The regulatory role of *BTNL2* was demonstrated in T cell activation and homeostasis as a costimulatory molecule during HLA class II antigen presentation to T cell receptors (Konno et al., 2009). Further, variants identified from the conditional analysis in White participants, namely, *rs116062523* and *rs7754362*, are eQTLs associated with the expression of *MICA/MICB* and *BRD2*, respectively. *MICA/MICB* and *BRD2* influence the activation and function of natural killer (NK) cells as a potential ligand and through the recognition of epigenetic markers, respectively (Stephens, 2001; Kerscher et al., 2019). Depending on the environment, NK cells can directly and indirectly regulate both activation and inhibition of $CD4^+$ and $CD8^+$ T cells (Vivier et al., 2008; Pallmer and Oxenius, 2016). Recent studies have shown that like $CD4^+$ T helper 1 (T_H1) and T_H2 cells, some $CD8^+$ subtypes, including cytotoxic type 1 and type 2 T cells (T_C1 and T_C2), can also produce type 1 and 2 cytokines (Lourenço et al., 2016). Some studies have shown that T_C2 cells are more strongly associated with corticosteroid insensitivity, persistent airway eosinophilia, severe asthma, and atopic asthma than $CD4^+$ T_H2 cells (Gelfand and Hinks, 2019). For White participants, there was notable interaction and/or transition between innate and adaptive immunity involving NK cell-dependent $CD4^+$ and $CD8^+$ T cell responses. Similar variants (and/or interactions) have not been previously reported in studies of early-onset asthma in a United States-based cohort stratified by race/ethnicity.

For Black participants, the HLA-DRB1*03:01 and HLA-DQB1*02:01 alleles were significantly associated with late-onset asthma. Most HLA II-associated autoimmune diseases are accounted for by HLA-DR2DQ6, HLA-DR4DQ8, and HLA-DR3DQ2 haplotypes, and HLA II alleles are much more diverse or polymorphic in African populations compared to White European populations (Mangalam et al., 2013). The HLA analysis results suggest that for Black late-onset asthma cases, T_H2 and T helper 17 (T_H17) cell-mediated pathways may play a role (Figure 6) (Harb et al., 2021). When inflammation occurs due to the presence of an HLA antibody, cells expressing HLA-DQ and HLA-DR activate $CD4^+$ T helper cells and promote the conversion of functional T regulatory cells (T_{reg}) to T_H17 T cells (Mangalam et al., 2013; Lion et al., 2016). Moreover, the plasticity of T_H17 cells allows the conversion of human T_H17 T cells into cells that produce both T_H2 and T_H17-related cytokines, including IL-4 and IL-17A, respectively (Cosmi et al., 2011). An intron variant, *rs17198965*, approximately 36,316 bp away from the

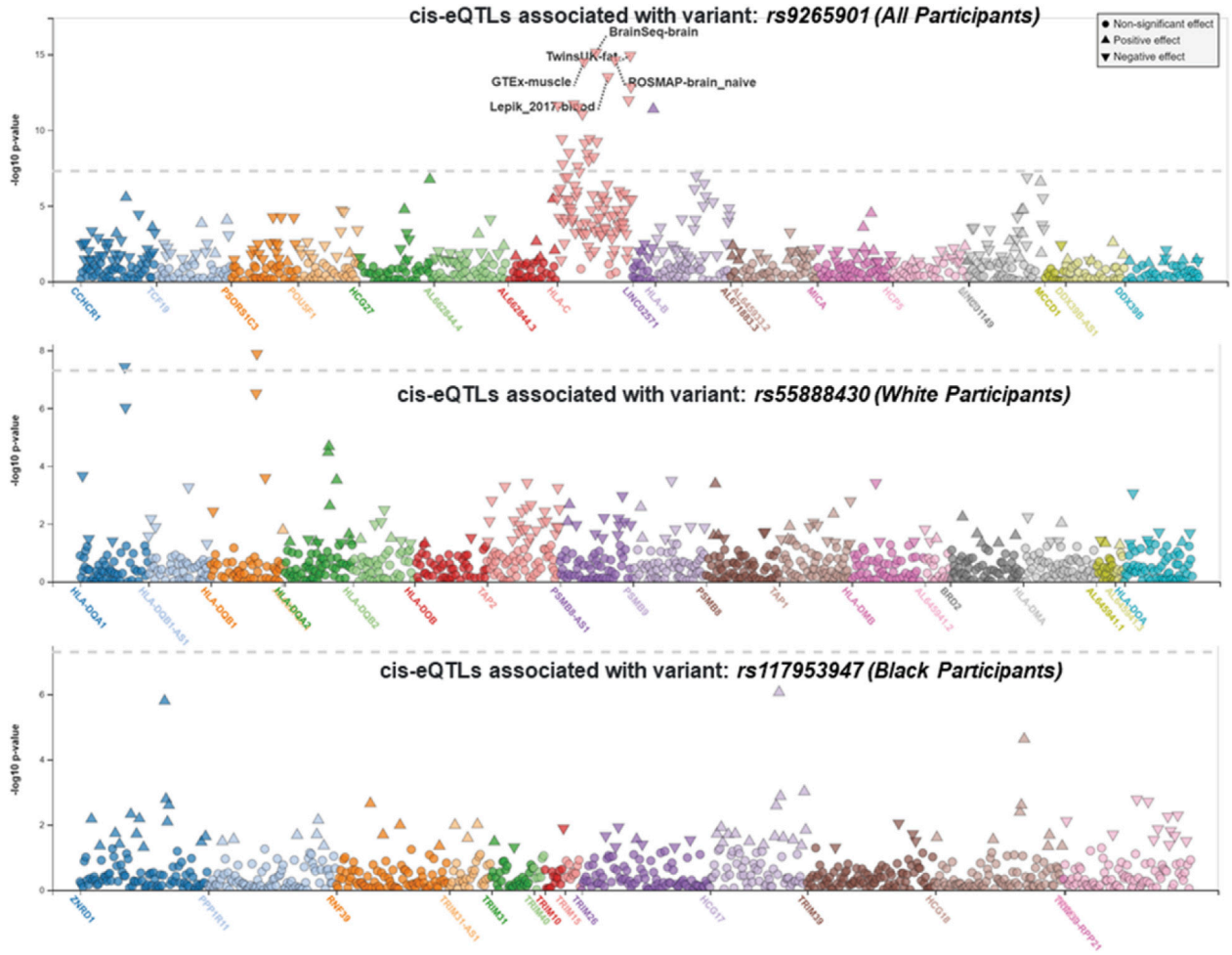


FIGURE 4
LocusZoom (RRID:SCR_009257) plots displaying cis-eQTLs. The lead SNPs for all participants (rs9265901), White participants (rs55888430), and Black participants (rs117953947) were weakly correlated with cis-eQTLs for multiple genes in multiple tissues (from top to bottom, respectively). The gray dotted line represents the significance threshold.

canonical transcription start site of HLA-C, had a strong promoter-like signature in lung tissue, bronchial epithelial cells, $CD8^+$ T cells, $CD4^+$ and T_H17 T cells, and B lymphocytes (cCRE accession ID: EH38E3700474). Further, *rs17198965* resides in the binding motif for the transcription factor *EZH2*, which suppresses major MHC class I molecule expression and is crucial for initiating $CD4^+$ T cell-associated response upon viral infection (Karantanos et al., 2016).

It is well-established that T_H2 and T_H17 cells mediate airway inflammation. The results of our conditional analyses are consistent with this and further show that the selected variant rs17422797 is associated with increased numbers of both eosinophils and neutrophils (βeta = 0.04, $p = 1.43 \times 10^{-20}$; βeta = 0.03, $p = 3.65 \times 10^{-18}$, respectively) (Broad Institute, 2021). Mechanistically, exposure to allergens and particulate matter promotes T_H2 and T_H17 T cell differentiation in a manner dependent on the interaction of *NOTCH4* receptors on T_{reg} cells with the Notch receptor ligand, *Jagged 1,* on antigen-presenting cells (Xia et al., 2018; Harb et al., 2021). Previous work has found associations in African Americans carrying *NOTCH4*-associated variants with sarcoidosis, which involves the formation of inflammatory cells (granulomas) in the body, primarily in the lungs (Adrianto et al., 2012).

African American individuals with asthma have an increased burden of symptoms and increased morbidity and mortality compared to White individuals with asthma and display resistance to inhaled corticosteroid treatment (Barnes et al., 2019). T_H17 is also linked to a difficult-to-treat, steroid-insensitive asthma phenotype (Chesné et al., 2014) and has been associated with difficult-to-control asthma in inner-city African American children and atopic dermatitis in African American individuals (Brown et al., 2017). Our findings propose candidate pathways that may be involved in the etiology of asthma in this population, related to population-specific allele frequencies and potential gene-gene and gene-environment interactions.

For HLA-DRB, HLA-B, and HLA-C, we examined the crystal structures with the highest resolution to determine the presence of functional consequences associated with selected individual residues in some alleles. Residues L37 and L/V38 of HLA-DRB are located in the peptide-binding surface and are in contact with the bound peptide. One of the two residues is also in contact with a helix from HLA-DRA that

TABLE 2 Fine-mapping of the MHC region harboring race/ethnicity-specific signals using sequenced data. We evaluated the joint effects of multiple variants on late-onset asthma with stepwise regression models. Conditional analyses of race/ethnicity-specific lead SNPs identified potential causal SNPs in the MHC region for all, White, and Black participants.

All participants						
Variants	Position	Gene	Alleles	Freq	OR (95% CI)[a]	*p*-value
rs9265901	6:31345615	HLA-B	A|G		1.01 (1.00, 1.03)	0.05
rs59377618	6:32820360	HLA-DOB	T|C		1.02 (1.01, 1.04)	4.42×10^{-3}
6:33097139	6:33097139	HLA-DPA1	G|A		0.98 (0.97, 0.99)	4.62×10^{-3}
rs2894334	6:33330567	SMIM40	A|C		0.99 (0.98, 0.99)	0.01
rs5009448	6:29972711	HLA-A	T|C		1.01 (1.00, 1.02)	0.02
rs7767589	6:30402196	RPP21	C|T		1.07 (1.02, 1.13)	7.00×10^{-3}
rs73728546	6:31380112	MICA	C|T		1.05 (1.02, 1.09)	5.24×10^{-3}
rs4569	6:31670030	LY6G5B	C|T		0.98 (0.97, 0.99)	3.21×10^{-3}
6:33061419	6:33061419	HLA-DPB1	C|T		1.10 (1.05, 1.16)	2.23×10^{-4}
rs9268541	6:32416750	BTNL2	T|C		1.03 (1.00, 1.05)	0.02
rs62407970	6:32969217	HLA-DMA	G|A		1.03 (1.01, 1.05)	6.89×10^{-3}
White participants						
rs55888430	6:32817365	HLA-DOB	G|A		1.03 (0.99, 1.07)	0.06
rs5009448	6:29972711	HLA-A	T|C		1.02 (1.01, 1.03)	1.65×10^{-3}
rs116062523	6:31383028	MICA	T|C		1.08 (1.03, 1.14)	3.64×10^{-3}
rs7754362	6:33143251	HCG24, COL11A2	A|C		0.98 (0.97, 0.99)	2.95×10^{-4}
6:29800571	6:29800571		A|C		1.07 (1.01, 1.13)	0.02
6:29816791	6:29816791		A|T		1.05 (1.00, 1.10)	0.04
6:31932537	6:31932537	CFB	T|C		1.07 (1.04, 1.11)	2.91×10^{-5}
rs1632859	6:31002121	MUC22	G|A		1.01 (1.00, 1.03)	0.04
rs17208209	6:32227599	NOTCH4	G|A		1.04 (1.00, 1.08)	0.04
rs115337486	6:33387741	KIFC1	C|T		1.04 (1.00, 1.08)	0.04
Black participants						
rs117953947	6:30244862	TRIM26, HCG17	G|A		1.13 (0.98, 1.30)	0.09
rs150891754	6:28608725	ZBED9	A|G		1.27 (1.13, 1.44)	1.20×10^{-4}
rs17200414	6:31449359	MICB	C|T		1.21 (1.05, 1.38)	7.90×10^{-3}
rs200535118	6:31352792	HLA-B	G|T		1.19 (1.01, 1.40)	0.04
rs62408569	6:29406052	OR5V1, OR12D2	C|T		1.31 (1.13, 1.52)	4.04×10^{-4}
rs181714996	6:33331782	SMIM40	T|C		1.19 (1.07, 1.32)	1.05×10^{-3}
rs115571466	6:32998090	HLA-DOA	A|G		1.20 (1.07, 1.35)	1.71×10^{-3}
rs4713423	6:31034524	MUC22	G|C		1.04 (1.00, 1.08)	0.03
rs17198965	6:31308446	HLA-C	C|G		1.08 (1.02, 1.16)	0.01
rs73403122	6:32501344	HLA-DRB5	A|G		1.21 (1.07, 1.36)	1.92×10^{-3}
rs17422797	6:32297751	NOTCH4	C|T		1.26 (1.09, 1.44)	1.38×10^{-3}
rs79735834	6:29517075	MAS1L	A|G		1.08 (1.02, 1.15)	0.01

(Continued on following page)

TABLE 2 (*Continued*) Fine-mapping of the MHC region harboring race/ethnicity-specific signals using sequenced data. We evaluated the joint effects of multiple variants on late-onset asthma with stepwise regression models. Conditional analyses of race/ethnicity-specific lead SNPs identified potential causal SNPs in the MHC region for all, White, and Black participants.

All participants						
Variants	Position	Gene	Alleles	Freq	OR (95% CI)[a]	*p*-value
rs187569734	6:31044172	MUC22	G|A		1.28 (1.11, 1.48)	6.82×10^{-4}

[a]Odds ratios represent the association between lead SNPs (*rs9265901*, *rs55888430*, *rs117953947*) and late-onset asthma, adjusted for covariates and additional SNPs in forward stepwise regression analyses. We adjusted all regression models for age, sex, and four ancestries.
Note: A full list of the variants identified from conditional analyses is available in the Supplementary Table S4.

TABLE 3 Associations and minor allele frequencies (MAF) of the leads SNPs across all strata.

			Pooled		EUR		AFR	
Lead SNPs	MAF	MAF	OR[a]	*p*-value	OR[a]	*p*-value	OR[a]	*p*-value
rs9265901	0.15 (G)	EUR: 21%	1.31	3.62×10^{-5}	1.66	6.91×10^{-4}	1.79	0.03
		AFR: 15%						
rs55888430	0.03 (A)	EUR: 4%	2.24	2.84×10^{-4}	3.05	8.85×10^{-6}	0.32	0.26
		AFR: 5%						
rs117953947	0.04 (A)	EUR: 2%	1.14	0.76	0.46	0.30	19.5	9.97×10^{-5}
		AFR: 1%						

[a]Odds ratios represent the association between lead SNPs (*rs9265901*, *rs55888430*, *rs117953947*) and late-onset asthma, adjusted for covariates and additional SNPs in forward stepwise regression analyses. We adjusted all regression models for age, sex, and four ancestries.

TABLE 4 HLA allele analysis results for all, White, and Black participants. Classical HLA alleles of class I and II genes in four-digit resolution were associated with increased and reduced odds of being a late-onset asthma case. We adjusted all regression models for age, sex, and four ancestries.

	Classical HLA allele analysis (4-digit)					
	All participants					
HLA-(gene)	* allele:protein	OR (95% CI)[a]	*p*-value	*n*	Freq[b]	Freq
HLA-B	**40:02*	3.18 (1.63 to 6.18)	$6.69 \times 10^{-4‡}$	95	0.03	0.02[c]
HLA-DRB1	**04:05*	6.94 (2.15 to 22.44)	$1.22 \times 10^{-3‡}$	18	0.005	0.0035[c]
White participants						
HLA-B	**40:02*	3.77 (1.91 to 7.44)	$1.28 \times 10^{-4§}$	82	0.03	0.02[c]
HLA-C	**04:01*	0.25 (0.09 to 0.70)	$7.97 \times 10^{-3§}$	325	0.12	0.13[c]
HLA-DRB1	**04:05*	8.19 (2.51 to 26.79)	$5.03 \times 10^{-4§}$	15	0.005	0.0035[c]
Black participants						
HLA-DRB1	**03:01*	4.00 (1.46 to 10.91)	6.71×10^{-3f}	29	0.05	0.07[d]
HLA-DQB1	**02:01*	2.25 (1.09 to 4.65)	2.91×10^{-2f}	132	0.25	0.22[d]

[a]Bi-allelic (presence or absence of the two given residues).
[b]HLA allele frequency computed from PEGS participants.
[c]HLA allele frequency computed from the United States population-Caucasian (n = 61,655)[35].
[d]HLA allele frequency computed from the United States population-African American (n = 2,411)[35].
HLA analysis: [‡]Pooled ancestry: significant p-value = 3.3×10^{-4}; suggestive p-value = 3.3×10^{-3}. [§]EUR ancestry: significant p-value = 2.0×10^{-3}; suggestive p-value = 2.0×10^{-2}. [f]AFR ancestry: significant p-value = 3.0×10^{-3}; suggestive p-value = 3.0×10^{-2}.

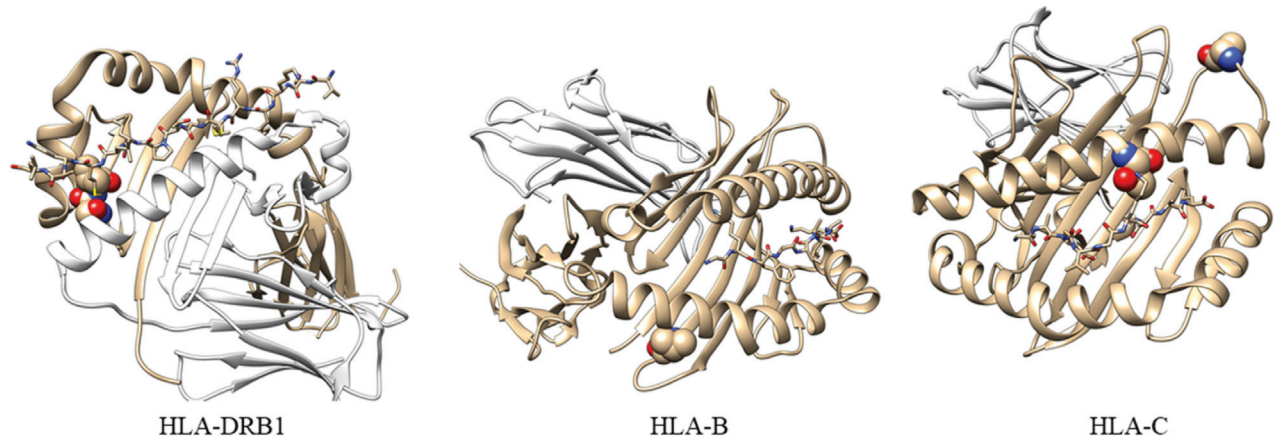


FIGURE 5
Analysis of amino acid residues showing their potential ability to bind antigens. **(A)** Atoms in residues 37 and 38 are shown as spheres. HLA-DRA (cyan), HLA-DRB1 (gray), M141 TCR? (wheat), and M141 TCRβ (yellow) are shown as ribbons. Residues 37 and 38 are located on the peptide-binding surface. One of the two residues is also in contact with a helix from HLA-DRA that comprises the complex. We created this figure using PDB ID (4x5w); **(B)** Atoms in residues 103 are shown as spheres. HLA-B (gray) and β-2-macroglobulin (wheat) are shown as ribbons. The small peptide-binding pocket (lines) is composed of three helices, one of which directly interacts with residue 103 within HLA-B. This residue is present at the end of a b-sheet while in contact with a few hydrophobic residues from the helix above and may be important for maintaining the structural scaffold and creating the binding surface We created this figure using PDB ID (1K5N); **(C)** Atoms in residues 73 and 90 are shown as spheres. HLA-C (gray) and β-2-macroglobulin (wheat) are shown as ribbons. The small binding peptide is in direct contact with residue 73, and residue 90 is located in a loop region just after the helix present on the peptide-binding surface. We created this figure using PDB ID (6JT0).

comprises the HLA complex. In HLA-B, L/V103 is an anchor residue of the helix that comprises the peptide-binding cavity and contributes to the stability of the hydrophobic patch composed of residues L109, V165, L168, and L172. Of the two residues selected from HLA-C, T73 is in direct contact with the peptide-binding surface while residue A90 is located in a loop region adjacent to the helix on the peptide-binding surface.

4.2 Strengths and limitations

The current study has several strengths. First, the collection of next-generation WGS data as opposed to genome-wide microarray-based ChIP-chip data enables access to comprehensive information at an allelic level across the diverse MHC locus. Second, enabled by the ancestry diversity in the PEGS cohort, we took advantage of natural differences in genomic LD across diverse populations to conduct transethnic gene mapping, enabling the prioritization of candidate genes, fine-mapping of functional variants, and potential identification of SNPs associated with disease risk in admixed populations.

Despite these strengths, the results must be interpreted in the context of important limitations. First, the importance of the sociocultural context of the varying results for Black and White subgroups cannot be understated. Race is a social construct correlated with health, economic, and exposure disparities. While socioeconomic status and other aspects of health disparities are important risk factors for many diseases, they are both a component and confounder of genetic associations and require further investigation (Fiscella and Williams, 2004). While genes and their variants are involved in the genetic etiology of asthma, regardless of race/ethnicity, the frequency of variants differs across strata, thus yielding varying results for significant associations. Additionally, the role of the environment, including structural racism, is crucially important, and gene-gene and gene-environment interactions may contribute to differing associations. Additionally, the small sample sizes of some of the stratified analyses may have affected statistical power, so the results should be interpreted with appropriate caution. Finally, we defined asthma cases using information on self-reported physician diagnoses. Given that asthma can be both over- and underdiagnosed (Aaron et al., 2018), this represents a potential confounder.

5 Conclusion

In this study, we explored the pathological role of genetic variants within the MHC region in late-onset asthma due to their critical importance in immune function/response and highly polymorphic nature. Our results suggest interesting parallels and contrasts in the genes and pathways that are associated with asthma dependent on race/ethnicity. The association of HLA II-associated genes and T_H2-related genes with asthma regardless of race/ethnicity may support a basic "minimum requirement" immune component to asthma that comprises increased susceptibility to sensitization and T_H2 activation. Various polymorphisms, depending on ancestry, may reflect stochastic genetic events (such as different allele frequencies) in populations but mechanistically induce similar effects. In contrast, in our cohort, we found a strong association between T_H17 pathways and late-onset asthma in Black participants. Our findings have important clinical implications because knowledge of race/ethnicity-specific

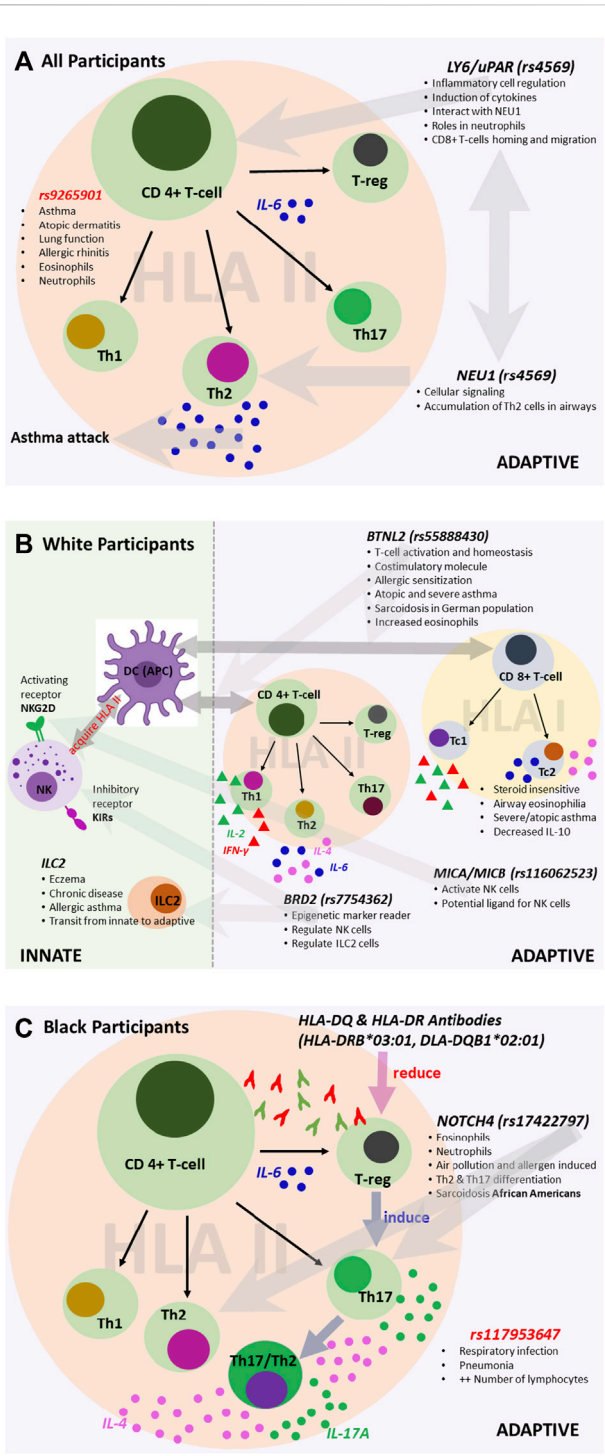


FIGURE 6
Variants associated with the immune system for **(A)** all participants, **(B)** White participants, and **(C)** Black participants. We reconstructed each pathway based on findings from the literature and sources that include the UK Biobank and Finnish Biobank. The figures illustrate the potential roles of the associated variants.

pathways for complex diseases such as late-onset asthma can help further dissect the etiology of the trait, which can ultimately assist with the development of precise, targeted treatments.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://gtexportal.org/home/ and https://www.ukbiobank.ac.uk/.

Author contributions

EL, JH, and AM-R developed the study design and concept. EL, AB, DF, and JH acquired the data used in the analysis. EL, WC, LP, JM, FM, MF, DC, PK, HN, SG, JM, JH, FA, and CS analyzed and interpreted the data. EL, WC, and LP performed statistical analysis. EL drafted the manuscript. WC, AB, LP, JM, FM, MF, DC, PK, HN, SG, JM, JH, FA, CS, DF, JH, and AM-R critically revised the manuscript. AM-R provided supervision. AM-R is responsible for the content of the manuscript, including the data and analysis. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the Intramural Research Program of the National Institutes of Health, National Institute of Environmental Health Sciences.

Acknowledgments

We would like to thank PEGS participants for their contributions. We would like to thank Hannah Collins Cakar for assistance with manuscript preparation.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1173676/full#supplementary-material

References

Aaron, S. D., Boulet, L. P., Reddel, H. K., and Gershon, A. S. (2018). Underdiagnosis and overdiagnosis of asthma. *Am. J. Respir. Crit. Care Med.* 198 (8), 1012–1020. doi:10.1164/rccm.201804-0682CI

Adrianto, I., Lin, C. P., Hale, J. J., Levin, A. M., Datta, I., Parker, R., et al. (2012). Genome-wide association study of African and European Americans implicates multiple shared and ethnic specific loci in sarcoidosis susceptibility. *PLoS One* 7 (8), e43907. doi:10.1371/journal.pone.0043907

American Thoracic Society (2018). *Asthma costs the U.S. economy more than $80 billion per year*. American Thoracic Society, Available: https://www.thoracic.org/about/newsroom/press-releases/journal/2018/asthma-costs-the-us-economy-more-than-80-billion-per-year.php (Accessed January 17, 2023).

Anderson, G. P. (2008). Endotyping asthma: New insights into key pathogenic mechanisms in a complex, heterogeneous disease. *Lancet* 372 (9643), 1107–1119. doi:10.1016/s0140-6736(08)61452-x

Barnes, P. J., Szefler, S. J., Reddel, H. K., and Chipps, B. E. (2019). Symptoms and perception of airway obstruction in asthmatic patients: Clinical implications for use of reliever medications. *J. Allergy Clin. Immunol.* 144 (5), 1180–1186. doi:10.1016/j.jaci.2019.06.040

Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 22 (9), 1790–1797. doi:10.1101/gr.137323.112

Broad Institute (2021). *Picard tools*. Broad Institute. Available: https://broadinstitute.github.io/picard/(Accessed).

Brown, K. R., Krouse, R. Z., Calatroni, A., Visness, C. M., Sivaprasad, U., Kercsmar, C. M., et al. (2017). Endotypes of difficult-to-control asthma in inner-city African American children. *PLoS One* 12 (7), e0180778. doi:10.1371/journal.pone.0180778

Carithers, L. J., Ardlie, K., Barcus, M., Branton, P. A., Britton, A., Buia, S. A., et al. (2015). A novel approach to high-quality postmortem tissue procurement: The GTEx Project. *Biopreserv. Biobank* 13 (5), 311–319. doi:10.1089/bio.2015.0032

Carvalho-Silva, D., Pierleoni, A., Pignatelli, M., Ong, C., Fumis, L., Karamanis, N., et al. (2019). Open targets platform: New developments and updates two years on. *Nucleic Acids Res.* 47 (D1), D1056–D1065. doi:10.1093/nar/gky1133

Centers for Disease Control (2022). *Most recent national asthma data*. Centers for Disease Control and Prevention. Available: https://www.cdc.gov/asthma/most_recent_national_asthma_data.htm (Accessed March 11, 2022).

Chesné, J., Braza, F., Mahay, G., Brouard, S., Aronica, M., and Magnan, A. (2014). IL-17 in severe asthma. Where do we stand? *Am. J. Respir. Crit. Care Med.* 190 (10), 1094–1101. doi:10.1164/rccm.201405-0859PP

Choi, W., Luo, Y., Raychaudhuri, S., and Han, B. (2021). Hatk: HLA analysis toolkit. *Bioinformatics* 37 (3), 416–418. doi:10.1093/bioinformatics/btaa684

Clay, S. M., Schoettler, N., Goldstein, A. M., Carbonetto, P., Dapas, M., Altman, M. C., et al. (2022). Fine-mapping studies distinguish genetic risks for childhood- and adult-onset asthma in the HLA region. *Genome Med.* 14 (1), 55. doi:10.1186/s13073-022-01058-2

Cook, S., Choi, W., Lim, H., Luo, Y., Kim, K., Jia, X., et al. (2021). Accurate imputation of human leukocyte antigens with CookHLA. *Nat. Commun.* 12 (1), 1264. doi:10.1038/s41467-021-21541-5

Cosmi, L., Liotta, F., Maggi, E., Romagnani, S., and Annunziato, F. (2011). Th17 cells: New players in asthma pathogenesis. *Allergy* 66 (8), 989–998. doi:10.1111/j.1398-9995.2011.02576.x

Daya, M., Cox, C., Acevedo, N., Boorgula, M. P., Campbell, M., Chavan, S., et al. (2021). Multiethnic genome-wide and HLA association study of total serum IgE level. *J. Allergy Clin. Immunol.* 148 (6), 1589–1595. doi:10.1016/j.jaci.2021.09.011

Fiscella, K., and Williams, D. R. (2004). Health disparities based on socioeconomic inequities: Implications for urban health care. *Acad. Med.* 79 (12), 1139–1147. doi:10.1097/00001888-200412000-00004

Galanter, J. M., Gignoux, C. R., Torgerson, D. G., Roth, L. A., Eng, C., Oh, S. S., et al. (2014). Genome-wide association study and admixture mapping identify different asthma-associated loci in latinos: The genes-environments & admixture in latino Americans study. *J. Allergy Clin. Immunol.* 134 (2), 295–305. doi:10.1016/j.jaci.2013.08.055

Gelfand, E. W., and Hinks, T. S. C. (2019). Is there a role for type 2 CD8(+) T cells in patients with steroid-resistant asthma? *J. Allergy Clin. Immunol.* 144 (3), 648–650. doi:10.1016/j.jaci.2019.07.022

Global Initiative for Asthma (2022). *Global strategy for asthma management and prevention (2022 update)*.

Habu, M., Koyama, H., Kishida, M., Kamino, M., Iijima, M., Fuchigami, T., et al. (2014). Ryk is essential for Wnt-5a-dependent invasiveness in human glioma. *J. Biochem.* 156 (1), 29–38. doi:10.1093/jb/mvu015

Harb, H., Stephen-Victor, E., Crestani, E., Benamar, M., Massoud, A., Cui, Y., et al. (2021). Author correction: A regulatory T cell notch4-GDF15 axis licenses tissue inflammation in asthma. *Nat. Immunol.* 22 (6), 794–795. doi:10.1038/s41590-021-00929-x

Hirano, T., and Matsunaga, K. (2018). Late-onset asthma: Current perspectives. *J. Asthma Allergy* 11, 19–27. doi:10.2147/jaa.S125948

Karantanos, T., Chistofides, A., Barhdan, K., Li, L., and Boussiotis, V. A. (2016). Regulation of T cell differentiation and function by EZH2. *Front. Immunol.* 7, 172. doi:10.3389/fimmu.2016.00172

Katoh, S., Maeda, S., Fukuoka, H., Wada, T., Moriya, S., Mori, A., et al. (2010). A crucial role of sialidase Neu1 in hyaluronan receptor function of CD44 in T helper type 2-mediated airway inflammation of murine acute asthmatic model. *Clin. Exp. Immunol.* 161 (2), 233–241. doi:10.1111/j.1365-2249.2010.04165.x

Kerscher, B., Barlow, J. L., Rana, B. M., Jolin, H. E., Gogoi, M., Bartholomew, M. A., et al. (2019). BET bromodomain inhibitor iBET151 impedes human ILC2 activation and prevents experimental allergic lung inflammation. *Front. Immunol.* 10, 678. doi:10.3389/fimmu.2019.00678

Kim, S. Y., Kim, E. G., Kim, M., Hong, J. Y., Kim, G. E., Jung, J. H., et al. (2021). Genome-wide association study identifies BTNL2 associated with atopic asthma in children. *Med. Baltim.* 100 (44), e27626. doi:10.1097/MD.0000000000027626

Konno, S., Takahashi, D., Hizawa, N., Hattori, T., Takahashi, A., Isada, A., et al. (2009). Genetic impact of a butyrophilin-like 2 (BTNL2) gene variation on specific IgE responsiveness to Dermatophagoides farinae (Der f) in Japanese. *Allergol. Int.* 58 (1), 29–35. doi:10.2332/allergolint.08-OA-0005

Lee, H., and Kingsford, C. (2018). Kourami: Graph-guided assembly for novel human leukocyte antigen allele discovery. *Genome Biol.* 19 (1), 16. doi:10.1186/s13059-018-1388-2

Lee, P. Y., Wang, J. X., Parisini, E., Dascher, C. C., and Nigrovic, P. A. (2013). Ly6 family proteins in neutrophil biology. *J. Leukoc. Biol.* 94 (4), 585–594. doi:10.1189/jlb.0113014

Lee, E. Y., Mak, A. C. Y., Hu, D., Sajuthi, S., White, M. J., Keys, K. L., et al. (2020). Whole-genome sequencing identifies novel functional loci associated with lung function in Puerto Rican youth. *Am. J. Respir. Crit. Care Med.* 202 (7), 962–972. doi:10.1164/rccm.202002-0351OC

Lee, E. Y., Akhtari, F., House, J. S., Simpson, R. J., Jr., Schmitt, C. P., Fargo, D. C., et al. (2022). Questionnaire-based exposome-wide association studies (ExWAS) reveal expected and novel risk factors associated with cardiovascular outcomes in the Personalized Environment and Genes Study. *Environ. Res.* 212, 113463. doi:10.1016/j.envres.2022.113463

Li, X., Howard, T. D., Zheng, S. L., Haselkorn, T., Peters, S. P., Meyers, D. A., et al. (2010). Genome-wide association study of asthma identifies RAD50-IL13 and HLA-DR/DQ regions. *J. Allergy Clin. Immunol.* 125 (2), 328–335. doi:10.1016/j.jaci.2009.11.018

Lion, J., Taflin, C., Cross, A. R., Robledo-Sarmiento, M., Mariotto, E., Savenay, A., et al. (2016). HLA class II antibody activation of endothelial cells promotes Th17 and disrupts regulatory T lymphocyte expansion. *Am. J. Transpl.* 16 (5), 1408–1420. doi:10.1111/ajt.13644

Lötvall, J., Akdis, C. A., Bacharier, L. B., Bjermer, L., Casale, T. B., Custovic, A., et al. (2011). Asthma endotypes: A new approach to classification of disease entities within the asthma syndrome. *J. Allergy Clin. Immunol.* 127 (2), 355–360. doi:10.1016/j.jaci.2010.11.037

Loughner, C. L., Bruford, E. A., McAndrews, M. S., Delp, E. E., Swamynathan, S., and Swamynathan, S. K. (2016). Organization, evolution and functions of the human and mouse Ly6/uPAR family genes. *Hum. Genomics* 10, 10. doi:10.1186/s40246-016-0074-2

Lourenço, O., Fonseca, A. M., and Taborda-Barata, L. (2016). Human CD8+ T cells in asthma: Possible pathways and roles for NK-like subtypes. *Front. Immunol.* 7, 638. doi:10.3389/fimmu.2016.00638

Luo, Y., Hitz, B. C., Gabdank, I., Hilton, J. A., Kagda, M. S., Lam, B., et al. (2020). New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res.* 48 (D1), D882–d889. doi:10.1093/nar/gkz1062

Mangalam, A. K., Taneja, V., and David, C. S. (2013). HLA class II molecules influence susceptibility versus protection in inflammatory diseases by determining the cytokine profile. *J. Immunol.* 190 (2), 513–518. doi:10.4049/jimmunol.1201891

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., and Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics* 26 (22), 2867–2873. doi:10.1093/bioinformatics/btq559

Martin, J. S., Xu, Z., Reiner, A. P., Mohlke, K. L., Sullivan, P., Ren, B., et al. (2017). HUGIn: Hi-C unifying genomic interrogator. *Bioinformatics* 33 (23), 3793–3795. doi:10.1093/bioinformatics/btx359

Pallmer, K., and Oxenius, A. (2016). Recognition and regulation of T cells by NK cells. *Front. Immunol.* 7, 251. doi:10.3389/fimmu.2016.00251

Pan-UK Biobank (2022). *Pan-UK Biobank: Pan-ancestry genetic analysis of the UK Biobank*. Available: https://pan.ukbb.broadinstitute.org/(Accessed March 11, 2022).

Pech, M., Weckmann, M., Konig, I. R., Franke, A., Heinsen, F. A., Oliver, B., et al. (2018). Rhinovirus infections change DNA methylation and mRNA expression in children with asthma. *PLoS One* 13 (11), e0205275. doi:10.1371/journal.pone.0205275

Poplin, R., Chang, P. C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., et al. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* 36 (10), 983–987. doi:10.1038/nbt.4235

Prugnolle, F., Manica, A., Charpentier, M., Guégan, J. F., Guernier, V., and Balloux, F. (2005). Pathogen-driven selection and worldwide HLA class I diversity. *Curr. Biol.* 15 (11), 1022–1027. doi:10.1016/j.cub.2005.04.050

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). Plink: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81 (3), 559–575. doi:10.1086/519795

Raj, A., Stephens, M., and Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197 (2), 573–589. doi:10.1534/genetics.114.164350

Rybicki, B. A., Walewski, J. L., Maliarik, M. J., Kian, H., Iannuzzi, M. C., and Group, A. R. (2005). The BTNL2 gene and sarcoidosis susceptibility in African Americans and Whites. *Am. J. Hum. Genet.* 77 (3), 491–499. doi:10.1086/444435

Stephens, H. A. (2001). MICA and MICB genes: Can the enigma of their polymorphism be resolved? *Trends Immunol.* 22 (7), 378–385. doi:10.1016/s1471-4906(01)01960-3

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., et al. (2015). UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 12 (3), e1001779. doi:10.1371/journal.pmed.1001779

Traherne, J. A., Horton, R., Roberts, A. N., Miretti, M. M., Hurles, M. E., Stewart, C. A., et al. (2006). Genetic analysis of completely sequenced disease-associated MHC haplotypes identifies shuffling of segments in recent human history. *PLoS Genet.* 2 (1), e9. doi:10.1371/journal.pgen.0020009

Vivier, E., Tomasello, E., Baratin, M., Walzer, T., and Ugolini, S. (2008). Functions of natural killer cells. *Nat. Immunol.* 9 (5), 503–510. doi:10.1038/ni1582

Xia, M., Harb, H., Saffari, A., Sioutas, C., and Chatila, T. A. (2018). A Jagged 1-Notch 4 molecular switch mediates airway inflammation induced by ultrafine particles. *J. Allergy Clin. Immunol.* 142 (4), 1243–1256. doi:10.1016/j.jaci.2018.03.009

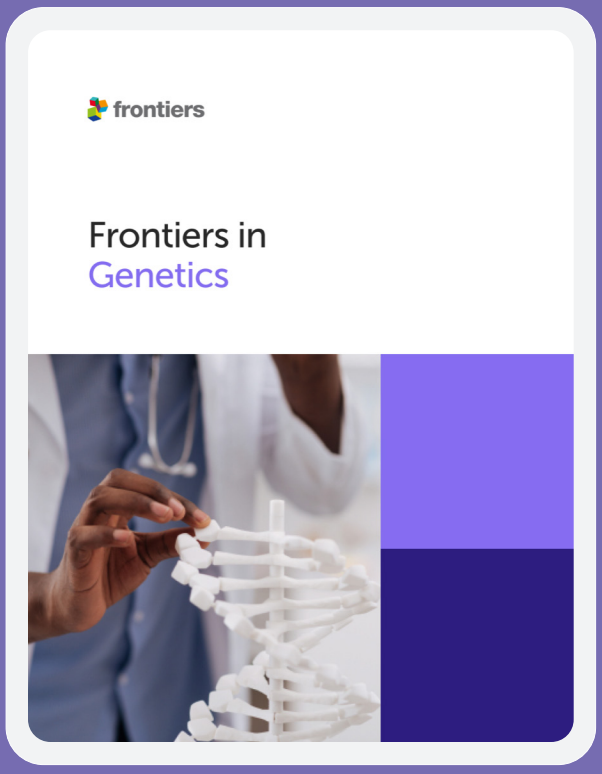
Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact