frontiers | Research Topics

Recent advances in understanding the genetics of immunological disorders

Edited by
Che Kang Lim, Jinqiao Sun and Hassan Abolhassani

Published in
Frontiers in Genetics
Frontiers in Pediatrics

ISSN 1664-8714
ISBN 978-2-8325-5778-5
DOI 10.3389/978-2-8325-5778-5

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

Recent advances in understanding the genetics of immunological disorders

Topic editors

Che Kang Lim — Department of Clinical Translational Research, Singapore General Hospital, Singapore

Jinqiao Sun — Fudan University, China

Hassan Abolhassani — Department of Medical Biochemistry and Biophysics, Karolinska Institutet (KI), Sweden

Citation

Lim, C. K., Sun, J., Abolhassani, H., eds. (2024). *Recent advances in understanding the genetics of immunological disorders*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-5778-5

Table of contents

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 02 December 2024
DOI 10.3389/fgene.2024.1530187

OPEN ACCESS

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

*CORRESPONDENCE
Che Kang Lim,
lim.che.kang@sgh.com.sg
Hassan Abolhassani,
hassan.abolhassani@ki.se
Jinqiao Sun,
jinqiaosun@fudan.edu.cn

RECEIVED 18 November 2024
ACCEPTED 19 November 2024
PUBLISHED 02 December 2024

CITATION
Lim CK, Abolhassani H and Sun J (2024) Editorial: Recent advances in understanding the genetics of immunological disorders. *Front. Genet.* 15:1530187. doi: 10.3389/fgene.2024.1530187

Editorial: Recent advances in understanding the genetics of immunological disorders

Che Kang Lim[1,2]*, Hassan Abolhassani[3,4]* and Jinqiao Sun[5]*

[1]Department of Clinical Translational Research, Singapore General Hospital, Singapore, Singapore, [2]Medicine Academic Clinical Programme (ACP), DUKE-NUS Medical School, Singapore, Singapore, [3]Division of Immunology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden, [4]Research Center for Immunodeficiencies, Pediatrics Center of Excellence, Children's Medical Center, Tehran University of Medical Sciences, Tehran, Iran, [5]Department of Clinical Immunology, National Children's Medical Center, Children's Hospital of Fudan University, Shanghai, China

KEYWORDS

immunological diseases, genetics, primary immumunodeficiencies, autoimmune diseases, allergy diseases, next-generation sequencing (NGS)

Editorial on the Research Topic
Recent advances in understanding the genetics of immunological disorders

Immunological disorders are conditions arising from immune system dysfunction, encompassing primary immunodeficiency syndromes, inborn errors of immunity, autoimmune diseases, asthma, allergies, lymphoproliferative and autoinflammatory syndromes. These disorders compromise the body's defense against pathogens and may cause exaggerated chronic inflammation responses and tissue damage. Owning clinical and genetic heterogeneity, immunological disorders exhibit a broad phenotypic spectrum, and distinct genotypes can produce overlapping clinical presentations. Additionally, the same mutation in immune-related genes within the same family can appear across multiple disorders, making accurate diagnosis and effective management complex and challenging.

In the past decade, remarkable advancements in genomics and sequencing technologies have revolutionized our understanding of candidate genes in immunological disorders. With new tools such as next-generation sequencing (NGS), allowing for in-depth genomic analysis, substantial progress has been made in identifying disease-causing genes associated with rare monogenic conditions, especially immunodeficiency syndromes (Vorsteveld et al., 2021; Conley and Casanova, 2014). These discoveries have yielded critical insights into the molecular mechanisms driving immune dysfunction, revealing how specific genetic mutations can impact immune pathways and trigger various immunological disorders. Importantly, the identification of these disease-causing genes has also paved the way for discovering novel biomarkers, which serve as valuable tools in clinical settings by enhancing diagnostic precision and aiding in the prediction of disease progression (Yazdanpanah and Rezaei, 2024; Peng and Kaviany, 2023). These biomarkers are instrumental in tailoring targeted therapeutic strategies, allowing for more effective and personalized treatment approaches. This genomic knowledge has not only enhanced our understanding of rare inborn errors of immunity but also laid the foundation for ongoing innovations in the diagnosis, management, and treatment of frequent complex immunological conditions. These findings shed insights into disease mechanisms and facilitated the identification of new biomarkers that can help in diagnosis, prognosis, and development of novel clinical therapeutic approaches.

This Research Topic presents eleven groundbreaking articles exploring the complex molecular genetics of immunological diseases. This compilation provides a comprehensive analysis of recent advancements in identifying the genetic and molecular mechanisms of immune-related disorders, including autoimmune diseases, allergies, and primary immunodeficiencies. It emphasizes the growing role of genetic insights, shared biological pathways, and biomarkers in revolutionizing clinical care. These studies collectively highlight the transformative potential of integrating innovative diagnostic techniques and precision medicine into the management of common immune disorders with high public health burdens.

In this Research Topic, significant findings span a range of conditions in autoimmune diseases, showcasing the interplay of genetic heterogeneity and shared pathways. For systemic sclerosis (SSc), Hanson et al. validated HLA Class II associations while uncovering a novel HLA Class I haplotype, HLA-B44:03-HLA-C16:01, linked to disease heterogeneity. This discovery, alongside the genetic interplay between killer cell immunoglobulin-like receptors (KIRs) and HLA ligands, offers critical insights into SSc's molecular mechanisms. Similarly, in rheumatoid arthritis (RA), Wen et al. identified 82 shared risk genes with other autoimmune conditions, such as multiple sclerosis (MS) and type 1 diabetes (T1D). These findings highlight overlapping disease mechanisms and the potential for cross-disease therapeutic strategies. Additionally, Ou et al. highlighted *TWIST1* gene as a key factor in ulcerative colitis (UC), while Chng et al. suggested that tear *S100A4* emerged as a biomarker for predicting thyroid eye disease (TED) in Graves' orbitopathy (GO). Fan et al. revealed that *CLEC16A* genetic variants, notably, the A allele of rs6498169 and the G allele of rs7200786, are associated with autoimmune diseases and also confer susceptibility to Parkinson's disease (PD) in Han Chinese. These findings emphasize the potential shared pathways between neurodegeneration and immune dysfunction. Finally, Yeo et al. reviewed articles on systemic lupus erythematosus (SLE) and discussed the implications of causal variants in both polygenic and monogenic forms of the disease. It suggests an age-based sequencing strategy to improve diagnostics and management, addressing genetic disparities and tailoring approaches to patient-specific factors.

Investigation in the field of allergy has also revealed important molecular mechanisms in the current Research Topic, particularly in allergic rhinitis (AR) and asthma. Sun et al. identified *ZNF667-AS1* as a key mediator in disease progression, driven by pollen-induced methylation and type 2 inflammatory pathways. The findings provide new avenues to mitigate disease progression in allergic disorders, reflecting the broader potential of genetic research to inform tailored treatments.

Some recent studies in the field of primary immunodeficiency or inborn errors of immunity featured the diagnostic utility of linking clinical phenotypes to underlying genetic defects. In common variable immunodeficiency (CVID), Cunningham-Rundles et al. examined a cohort of 405 patients and highlighted the effectiveness of phenotype-guided genetic diagnostics, particularly in cases involving autoinflammation or autoimmunity. This approach helps pinpoint the specific genetic mutations responsible for these predominantly antibody-deficient conditions, improving the accuracy of diagnoses, and paving the way for more targeted interventions.

Three case reports on rare disorders in the Research Topic further illustrate the importance of advanced genetic testing. Li et al. reported a child with COPA syndrome presented with interstitial lung disease and neuromyelitis optica spectrum disorder (NMOSD), underscoring the utility of genomic sequencing in diagnosing rare phenotypes. Deng et al. presented a case involving severe combined immunodeficiency (SCID) caused by compound heterozygous mutations in the *DCLRE1C* gene. Advanced genetic testing revealed a unique manifestation of rubella virus-induced cutaneous granulomas, which had not been previously associated with SCID. Additionally, Chen et al. identified novel variants in *TCRIG1* and *CLCN7* genes, contributing to a deeper understanding of genotype-phenotype correlations in osteopetrosis. These findings underscore the importance of further research into the genetic mechanisms and potential treatments for these complex immune-related disorders.

In conclusion, this Research Topic highlights the translational impact of genetic research on the diagnosis, treatment and prognosis estimation of immunological disorders. By uncovering novel genetic associations, unique biomarkers, and shared pathways, these studies provide a strong foundation for the development of targeted/personalized medicine. Integrating these insights into clinical practice holds significant potential to improve patient outcomes, especially in managing the complex and heterogeneous nature of immune conditions. As genetic research continues to evolve, it promises to unlock new opportunities for individualized care and therapeutic innovations. We highly believe that the compilation of publications within this Research Topic offers a valuable snapshot of the current focus and progress in the field of genetic research on human immunological disorders.

Author contributions

CL: Conceptualization, Writing–original draft, Writing–review and editing. HA: Conceptualization, Writing–review and editing. JS: Conceptualization, Writing–review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The authors would like to thank all the contributing authors, peer reviewers, and the editorial staff for their significant contributions and great support in the development and implementation of this Research Topic.

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 author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Conley, M. E., and Casanova, J. L. (2014). Discovery of single-gene inborn errors of immunity by next generation sequencing. *Curr. Opin. Immunol.* 30, 17–23. doi:10.1016/j.coi.2014.05.004

Peng, X., and Kaviany, S. (2023). Approach to diagnosing inborn errors of immunity. *Rheumatic Dis. Clin. N. Am.* 49 (4), 731–739. doi:10.1016/j.rdc.2023.06.001

Vorsteveld, E. E., Hoischen, A., and van der Made, C. I. (2021). Next-generation sequencing in the field of primary immunodeficiencies: current yield, challenges, and future perspectives. *Clin. Rev. Allergy Immunol.* 61 (2), 212–225. doi:10.1007/s12016-021-08838-5

Yazdanpanah, N., and Rezaei, N. (2024). The multidisciplinary approach to diagnosing inborn errors of immunity: a comprehensive review of discipline-based manifestations. *Expert Rev. Clin. Immunol.* 20 (10), 1237–1259. doi:10.1080/1744666X.2024.2372335

frontiers in Genetics

ORIGINAL RESEARCH
published: 30 March 2022
doi: 10.3389/fgene.2022.856493

Autoimmune Disease Associated *CLEC16A* Variants Convey Risk of Parkinson's Disease in Han Chinese

Hui-Hui Fan[1,2†], Lei Cui[1†], Xiao-Xia Jiang[1], Ya-Dan Song[1], Shu-Shu Liu[1], Ke-Yun Wu[1], Hao-Jia Dong[1], Miao Mao[1], Begench Ovlyakulov[1], Hong-Mei Wu[1], Jian-Hong Zhu[1,2]* and Xiong Zhang[1,2]**

[1]Department of Preventive Medicine, Institute of Nutrition and Diseases, Wenzhou Medical University, Wenzhou, China, [2]Department of Neurology, Institute of Geriatric Neurology, the Second Affiliated Hospital and Yuying Children's Hospital, Wenzhou Medical University, Wenzhou, China

OPEN ACCESS

Edited by:
Corrado Romano,
University of Catania, Italy

Reviewed by:
Félix Javier Jiménez-Jiménez,
Hospital Universitario del Sureste, Spain
Paul Lasko,
McGill University, Canada

***Correspondence:**
Hong-Mei Wu
hmwu@wmu.edu.cn
Jian-Hong Zhu
jhzhu@wmu.edu.cn
Xiong Zhang
zhangxiong98@gmail.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: 17 January 2022
Accepted: 09 February 2022
Published: 30 March 2022

Citation:
Fan H-H, Cui L, Jiang X-X, Song Y-D, Liu S-S, Wu K-Y, Dong H-J, Mao M, Ovlyakulov B, Wu H-M, Zhu J-H and Zhang X (2022) Autoimmune Disease Associated CLEC16A Variants Convey Risk of Parkinson's Disease in Han Chinese. Front. Genet. 13:856493. doi: 10.3389/fgene.2022.856493

CLEC16A is a membrane-associated endosomal protein implicated in regulating autophagy and antigen presentation. Its genetic variants are broadly associated with multiple autoimmune diseases. Parkinson's disease (PD), which undergoes autophagy disruption and neuroinflammation, has been clinically observed, for an extensive amount of time, to be associated with autoimmune diseases. In this study, we aimed to understand whether the autoimmune disease associated *CLEC16A* variants pleiotropically modulate PD risk. Five of such *CLEC16A* variants, including rs6498169, rs12708716, rs12917716, rs7200786, and rs2903692, were selected and analyzed in a Han Chinese cohort comprising 515 sporadic PD patients and 504 controls. Results showed that rs6498169 and rs7200786 were significantly associated with PD susceptibility (p = 0.005 and 0.004, respectively; recessive model, p = 0.002 and 0.001, respectively). Rs6498169 was also associated with the PD subtype of postural instability/gait difficulty (p = 0.002). Haplotype analysis showed that the AAG module in order of rs6498169, rs12708716, and rs2903692 was associated with the highest risk for PD (p = 0.0047, OR = 1.42, 95% CI = 1.11–1.82). Functional annotation analyses suggested that rs6498169 had high probability to affect transcription factor binding and target gene expression. In summary, the current study demonstrates that the autoimmune disease associated *CLEC16A* variants convey risk of PD in Han Chinese. Our findings suggest a pleiotropic role of *CLEC16A* and strengthen the link between PD and autoimmune diseases.

Keywords: CLEC16A, Parkinson's disease, autoimmune diseases, genetic variation, association

INTRODUCTION

CLEC16A (C-type lectin domain family 16, member A; Chr16; Gene ID: 23274; MIM: 611303), encoding a large protein of 1,053 amino acids with a putative C-type lectin-like domain, is a membrane-associated endosomal protein (Soleimanpour et al., 2014). *CLEC16A* genetic variations have been associated with multiple autoimmune diseases including multiple sclerosis, type 1 diabetes, Crohn's disease, Addison's disease, rheumatoid arthritis, and juvenile idiopathic arthritis (Vitale et al., 2002; Hakonarson et al., 2007; International Multiple Sclerosis Genetics et al., 2007; Marquez et al., 2009; Martinez et al., 2010; Skinningsrud et al., 2010; Nischwitz et al.,

2011; International Multiple Sclerosis Genetics et al., 2013). The broad association with autoimmune disorders suggests that *CLEC16A* may functionally link to autoimmunity by certain common pathogenic pathways. Indeed, CLEC16A has been implicated in regulating autophagy/mitophagy and antigen presentation. CLEC16A interacts with E3 ubiquitin ligase NRDP1 (neuregulin receptor degradation protein 1), controlling the volume of Parkin and its mastered mitophagy and thereby mediates murine β cell function and diabetogenesis (Soleimanpour et al., 2014). In addition, *Clec16a* knockdown in the nonobese diabetic mice protects against autoimmunity by altering the T cell selection, possibly through the inhibition of thymic epithelial cell autophagy (Schuster et al., 2015). *Clec16a* mutant mice generated by gene-trap insertion develop neurodegeneration featured by impaired motor behaviors, loss of Purkinje cells, and abnormal autophagy homeostasis (Redmann et al., 2016). CLEC16A also plays a role in multiple sclerosis *via* controlling the late endosome biogenesis-mediated human leukocyte antigen class II antigen presentation (van Luijn et al., 2015).

Parkinson's disease (PD) is the second most common neurodegenerative disorder characterized by the loss of dopaminergic neurons in the substantia nigra pars compacta. The majority of PD is sporadic and its etiology is complex and remains largely unknown. Autophagy dysfunction is one of the key mechanisms involved in PD, leading to the abnormal degradation and aggregation of proteins such as α-synuclein (Michel et al., 2016). Mutations in *PRKN* (Parkin encoding gene) are the common cause of the early-onset PD (Kalia and Lang, 2015). Recent studies have also revealed the presence of α-synuclein-reactive T cells in PD patients and an association of its reactivity with preclinical and early PD, indicating a link of PD to autoimmunity (Sulzer et al., 2017; Lindestam Arlehamn et al., 2020).

Indeed, PD risk has been observed, for an extensive amount of time, as associated with autoimmune diseases. For instance, a Swedish epidemiological study showed that subsequent risks of PD are increased in patients with 6 autoimmune diseases, including amyotrophic lateral sclerosis, Graves's disease/hyperthyroidism, Hashimoto's disease/hypothyroidism, multiple sclerosis, pernicious anemia, and polymyalgia rheumatica (Li et al., 2012), whereas being a casual or causal association between PD risk and multiple sclerosis remains in debate (Pedemonte et al., 2013; Nielsen et al., 2014). Nonetheless, it is hypothesized that there are common genetic risk variants between PD and autoimmune diseases (Witoelar et al., 2017). Given the functional overlap of *CLEC16A* with PD pathogenic mechanisms and its broad association with autoimmune diseases, we aimed, in this study, to investigate whether the autoimmune disease associated *CLEC16A* variants convey risk of PD in a Han Chinese population.

MATERIALS AND METHODS

Subjects

A total of 1,019 subjects of Han Chinese ethnicity from eastern China were recruited in this study, comprising 515 sporadic PD patients (263 males and 252 females) and 504 controls (265 males and 239 females). The median age of the patients and controls was 66 (interquartile range: 60–73) and 66 (interquartile range: 57–72) years old, respectively. All PD patients were diagnosed by two movement disorder neurologists, according to the UK Parkinson's disease Society Brain Bank Criteria (Hughes et al., 1992). Patients with a family history of PD or with secondary and atypical parkinsonism were excluded. The controls had no neurological disorders according to medical history, physical, and laboratory examinations. All subjects participating in the study signed written informed consents. The study was performed under the approval No. LCKY 2020-66 by the Ethics Committee of The Second Affiliated Hospital and Yuying Children's Hospital, Wenzhou Medical University.

Selection of *CLEC16A* Risk Variants

Five autoimmune disease-associated risk variants of *CLEC16A* were selected, including rs6498169, rs12708716, rs7200786, rs2903692 and rs12917716. Among these, rs6498169, rs12708716 and rs7200786 were identified as susceptibility markers for multiple sclerosis (Vitale et al., 2002; International Multiple Sclerosis Genetics et al., 2007; International Multiple Sclerosis Genetics et al., 2013); rs2903692 and rs12917716 were found to be associated with type 1 diabetes (Hakonarson et al., 2007; Skinningsrud et al., 2010); rs6498169 was additionally associated with juvenile idiopathic arthritis and rheumatoid arthritis (Martinez et al., 2010; Skinningsrud et al., 2010); rs2903692 was additionally associated with Crohn's disease and multiple sclerosis (Marquez et al., 2009; Martinez et al., 2010); rs12917716 was additionally associated with Addison's disease, multiple sclerosis, and primary adrenal insufficiency (Skinningsrud et al., 2010; Nischwitz et al., 2011).

Genotyping

Genomic DNA was extracted from peripheral blood using TIANamp Genomic DNA kit (Tiangen, Beijing, China) according to the manufacturer's instruction. Four SNPs, including rs6498169, rs12708716, rs12917716, and rs7200786, were genotyped using SNaPshot at BGI Technology (Wuhan, China) as described previously (Zou et al., 2018). Rs2903692 was not successfully genotyped by SNaPshot, and therefore was genotyped by conventional PCR and sequencing. The PCR condition was initial denaturation at 95°C for 3 min, followed by 32 cycles of 95°C for 30 s, 60°C for 30 s, and 72°C for 60 s, and a final extension at 72°C for 5 min. The PCR products were sequenced at BGI Technology (Wuhan, China). All relevant primers were listed in **Supplementary Table S1**.

Statistical Analysis

Statistical analyses were carried out using the Statistical Package for Social Science program (SPSS for Windows, version 23.0). Hardy-Weinberg equilibrium in genotype distribution was assessed using $\chi 2$ test. Following Kolmogorov-Smirnov test for normality, Mann-Whitney U test was used to evaluate age difference. The $\chi 2$ test was also used to assess differences in

TABLE 1 | Genotype and allele frequencies of five *CLEC16A* variants in PD patients and controls.

Variant	Genotype, n (%)			*P*	Allele, n (%)		*P*	OR (95%CI)
rs6498169	GG	GA	AA	0.005[a]	G	A	0.042	1.199 (1.007–1.427)
Control	151 (30.0)	264 (52.4)	89 (17.7)		566 (56.2)	442 (43.8)		
PD	150 (29.1)	232 (45.0)	133 (25.8)		532 (51.7)	498 (48.3)		
rs12708716	AA	AG	GG	0.430	A	G	0.288	1.116 (0.912–1.366)
Control	293 (58.1)	186 (36.9)	25 (5.0)		772 (76.6)	236 (23.4)		
PD	288 (55.9)	192 (37.3)	35 (6.8)		768 (74.6)	262 (25.4)		
rs12917716	GG	GC	CC	0.055	G	C	0.026	1.221 (1.024–1.456)
Control	173 (34.3)	258 (51.2)	73 (14.5)		604 (59.9)	404 (40.1)		
PD	154 (29.9)	259 (50.3)	102 (19.8)		567 (55.0)	463 (45.0)		
rs7200786	AA	AG	GG	0.004[a]	A	G	0.044	1.207 (1.005–1.448)
Control	220 (43.7)	239 (47.4)	45 (8.9)		679 (67.4)	329 (32.6)		
PD	216 (41.9)	218 (42.3)	81 (15.7)		650 (63.1)	380 (36.9)		
rs2903692	GG	GA	AA	0.731	G	A	0.732	1.036 (0.846–1.268)
Control	289 (57.3)	188 (37.3)	27 (5.4)		766 (76.0)	242 (24.0)		
PD	286 (55.5)	204 (39.6)	25 (4.9)		776 (75.3)	254 (24.7)		

[a]P < 0.01. CI, confidence interval; OR, odds ratio; PD, Parkinson's disease.

gender, genotype, allele, and haplotype frequencies between the PD patients and controls. The haplotype construction and association analysis were performed using SNPStats Online Version (https://www.snpstats.net/start.htm) (Sole et al., 2006). A backward elimination method as reported previously was used to identify the highest-risk haplotype for PD (Francis et al., 2007). Statistical power was calculated by the QUANTO version 1.2.4. A p value < 0.05 was considered statistically significant unless otherwise indicated.

Function Annotation

Function annotations of the variants were obtained from HaploReg v4.1 (http://pubs.broadinstitute.org/mammals/haploreg/haploreg.php) and RegulomeDB (http://www.regulomedb.org/). HaploReg was used to annotate potential causal links to disease pathogenesis (Ward and Kellis, 2016). RegulomeDB was used to annotate known and predicted genetic variations in regulatory elements in intergenic regions of the human genome. A score ranging from 1 to 6 was provided to indicate the potential function and the lower score indicates a higher probability that a variant affects binding and gene expression (Boyle et al., 2012). The expression quantitative trait locus (eQTL) and splicing quantitative trait locus (sQTL) of the variants were analyzed by the GTEx Portal (https://gtexportal.org/).

TABLE 2 | Genetic model analysis of rs6498169 and rs7200786.

Model	Genotype	*P*	OR (95% CI)
rs6498169			
Recessive	GG + GA *vs.* AA	0.002[a]	1.629 (1.203–2.205)
Dominant	GG *vs.* GA + AA	0.728	1.049 (0.801–1.374)
Additive	GG *vs.* GA *vs.* AA	0.039	1.200 (1.009–1.428)
rs7200786			
Recessive	AA + AG *vs.* GG	0.001[a]	1.895 (1.285–2.795)
Dominant	AA *vs.* AG + GG	0.569	1.075 (0.838–1.379)
Additive	AA *vs.* AG *vs.* GG	0.046	1.205 (1.004–1.446)

[a]P < 0.01. CI, confidence interval; OR, odds ratio; PD, Parkinson's disease.

TABLE 3 | Haplotype analysis of the *CLEC16A* variants in PD patients and controls.

Haplotype[a]	Control, n (%)	PD, n (%)	*P*	OR (95% CI)
GAGAG	453.5 (44.99)	399.7 (38.81)	—	1.00
AGCGA	182.8 (18.13)	197.0 (19.13)	0.12	1.22 (0.95–1.56)
AAGAG	93.3 (9.26)	120.9 (11.74)	0.02[b]	1.47 (1.06–2.04)
AACGG	82.2 (8.15)	98.3 (9.54)	0.03[b]	1.47 (1.04–2.07)
GACAG	72.2 (7.16)	75.8 (7.38)	0.34	1.20 (0.82–1.74)
Total	1008.0 (100)	1030.0 (100)	0.034[b]	—

[a]Haplotype alleles were in the order of rs6498169, rs12708716, rs12917716, rs7200786, and rs2903692. Haplotypes with frequency <3% in both PD patients and controls were excluded from the analysis.
[b]P < 0.05. CI, confidence interval; OR, odds ratio; PD, Parkinson's disease.

RESULTS

Association of the CLEC16A Variants with PD Susceptibility

Genotype distributions of the 5 variants in controls were in accordance with Hardy-Weinberg equilibrium (p > 0.05). The PD cases and controls were comparable in both gender and age (p > 0.05). Difference was considered only after Bonferroni correction (threshold for significance = 0.01). Results showed that significant difference between PD cases and control cases was found only in genotypes of rs6498169 and rs7200786 (p = 0.005 and p = 0.004, respectively; **Table 1**). No difference was found within rs12708716, rs12917716, and rs2903692.

We further analyzed the rs6498169 and rs7200786 by three genetic models (additive, dominant, and recessive). Both SNPs were significantly associated with PD in the recessive model (p = 0.002, OR = 1.629, 95% CI = 1.203–2.205 for rs6498169; p = 0.001, OR = 1.895, 95% CI = 1.285–2.795 for rs7200786). The risk genotypes for PD were the AA of rs6498169 and the GG of rs7200786 (**Table 2**). The statistical power was 92% and

TABLE 4 | The highest-risk haplotype analysis of *CLEC16A* in association with PD.

Variants, n	Haplotype[a]	Control, n (%)	PD, n (%)	*P*	OR (95%CI)
5	AAGAG	93.3 (9.26)	120.9 (11.74)	0.02	1.47 (1.06–2.04)
4	AAG-G	96.7 (9.59)	129.4 (12.56)	0.011	1.54 (1.11–2.15)
3	AA--G	185.1 (18.36)	233.1 (22.63)	0.0047	1.42 (1.11–1.82)
2	A---G	218.0 (21.63)	264.8 (25.71)	0.017	1.32 (1.05–1.65)

[a]*Haplotype alleles were in the order of rs6498169, rs12708716, rs12917716, rs7200786, and rs2903692. Haplotypes with frequency <3% in both PD patients and controls were excluded from the analysis. A hyphen indicates the eliminated variant at that position.*
CI, confidence interval; OR, odds ratio; PD, Parkinson's disease.

TABLE 5 | Function annotations of the PD-associated *CLEC16A* variants.

Tools	rs6498169	rs7200786	
HaploReg			
Enhancer histone marks[a]	13 tissues	13 tissues	
DNase hypersensitivity[a]	7 tissues	3 tissues	
Proteins bound	GATA2	—	
Motifs changed	HES1	CCNT2; EBF	
RegulomeDB			
Score	1b	4	
GTEx (eQTL)			
Genes affected	*RMI2*	*CLEC16A*	*RMI2*
Tissues affected[b]	Cerebellum; cortex; frontal cortex; hypothalamus; nucleus accumbens; putamen; spinal cord	Amygdala; cerebellum; cerebellar hemisphere	Anterior cingulate; cerebellum; cortex; nucleus accumbens; putamen; spinal cord
GTEx (sQTL)			
Genes affected	*CLEC16A*	*CLEC16A*	
Tissues affected[c]	Testis	Testis	

[a]*Tissues were detailed in* ***Supplementary Table S3****.*
[b]*Brain tissues were listed herein. Full tissue lists were shown in* ***Supplementary Figures S1–S3****.*
[c]*Only testis data were available. Details were shown in* ***Supplementary Figure S4****.*
eQTL, expression quantitative trait locus; PD, Parkinson's disease; sQTL, splicing quantitative trait locus.

96%, respectively, for the recessive model of rs6498169 and rs7200786.

We also extracted a total of 265 PD patients recorded with PD subtypes to understand their association with the five *CLEC16A* SNPs. The subtypes were classified into postural instability/gait difficulty (PIGD), tremor dominant, and indeterminate. Results showed that the genotype distribution of rs6498169 was significantly different (p = 0.002) between the PIGD patients and the controls. No other difference was found in the variants between the PD subtypes and the controls (**Supplementary Table S2**).

Haplotype Analyses of the CLEC16A Variants

We analyzed whether haplotypes of these *CLEC16A* variants were associated with PD. Haplotypes were constructed in the following order: rs6498169, rs12708716, rs12917716, rs7200786, and rs2903692. As shown in **Table 3**, five haplotypes were listed as those with a frequency < 3.0% in both PD patients and controls were excluded. A significant difference (p = 0.034) in overall haplotype distribution was observed between the PD patients and controls. Two haplotypes, AAGAG (p = 0.02, OR = 1.47, 95% CI = 1.06–2.04) and AACGG (p = 0.03, OR = 1.47, 95% CI = 1.04–2.07), showed a significant difference between the cases and controls.

To characterize the highest-risk haplotype of *CLEC16A* towards PD, a backward elimination model was employed. Results showed that the strongest PD-associated haplotype was AAG in the order of rs6498169, rs12708716, and rs2903692 from the best 3-variant model (p = 0.0047, OR = 1.42, 95% CI = 1.11–1.82; **Table 4**).

Function Annotations of the PD-Associated CLEC16A Variants

Function annotations of the PD-associated variants were performed using the HaploReg, RegulomeDB, and GTEx. Based on the HaploRreg (**Tables 5** and **Supplementary Table S3**), rs6498169 was located within enhancer histone marks in 13 tissues, DNase hypersensitivity regions in 7 tissues, and the region of GATA2 (GATA binding protein 2) binding site, and was predicted to significantly alter the binding motif of the HES1 (hes family bHLH transcription factor 1) transcription factor. Rs7200786 was located within enhancer histone marks in 13 tissues, and DNase hypersensitivity regions in 3 tissues and was predicted to significantly change the binding motifs of the CCNT2 (cyclin T2) and EBF (early B cell factor) transcription factors.

In RegulomeDB (**Table 5**), where the score <3 indicates a relatively high possibility of potential regulatory function (Luciano et al., 2018), rs6498169 was predicted with a score of 1b, representing that this variant is highly likely to affect transcription factor binding, certain motifs, DNase footprint and peaks, and potentially affect the expression of target genes. The score of rs7200786 was 4, representing that this variant may affect transcription factor binding and DNase peaks.

By using GTEx (**Table 5**; **Supplementary Figures S1–S3**), rs6498169 was suggested to be significant eQTL of the *RMI2* (RecQ mediated genome instability 2) expression in brain tissues such as cortex ($p = 0.0015$), cerebellum ($p = 0.008$) and nucleus accumbens ($p = 0.008$). Rs7200786 was as significant eQTL of the *RMI2* expression in brain tissues such as nucleus accumbens ($p = 0.00071$), cortex ($p = 0.0028$) and putamen ($p = 0.0033$), as well as the *CLEC16A* expression in brain tissues such as cerebellar hemisphere ($p = 0.0012$) and amygdala ($p = 0.008$). However, both rs6498169 and rs7200786 were not associated with the expression of *RMI2* and *CLEC16A* in the substantia nigra. Results of the sQTL analysis showed that the risk alleles of both rs6498169 and rs7200786 (A and G, respectively) were significantly associated with an increased intron-excision ratio, which potentially leads to elevation of certain variant expression of *CLEC16A* (**Supplementary Figure S4**).

DISCUSSION

Immunity disturbance is increasingly considered to be important in PD pathogenesis. The clinical observations of risk association between PD and autoimmune diseases indicate the possibility of having common genetic risk loci in between. By studying the five autoimmune disease associated *CLEC16A* variants in a Chinese cohort, we demonstrate that *CLEC16A* is pleiotropic for modulating PD risk. Two loci, rs6498169 and rs7200786, are recessively associated with PD susceptibility.

CLEC16A is involved in regulation of autophagy, T cell selection, antigen presentation, and neurodegeneration as suggested earlier (Soleimanpour et al., 2014; Schuster et al., 2015; van Luijn et al., 2015; Redmann et al., 2016). Autophagy plays a role, not only for substance clearance and recycling, but also in the presentation of antigenic peptides to the receptor of T cells in the context of antigen-presenting cells and major histocompatibility complex class II (Bonam and Muller, 2020). While disruption of the autophagy pathway serves as one of the key mechanisms in PD (Michel et al., 2016), PD patients are indeed observed with dysregulated innate and adaptive immune responses, particularly in those carrying autophagy-related gene mutations such as in Parkin and LRRK2 (leucine rich repeat kinase 2). Besides the previously mentioned α-synuclein reactive T cell responses (Sulzer et al., 2017; Lindestam Arlehamn et al., 2020), B cell-produced autoantibodies against α-synuclein antigen, GM1 gangliosides, and neuronal antigens are also shown to be elevated in blood and/or cerebral spinal fluid of PD patients (Zappia et al., 2002; van de Warrenburg et al., 2008; Scott et al., 2018). Hence, besides the identified genetic connection, *CLEC16A* may be functionally possible to participate in PD.

Our results show that rs6498169 and rs7200786 are associated with PD susceptibility. The modulation of rs7200786 on PD risk is consistent with a previous report in the Italian population (Strafella et al., 2021). The A of rs6498169 and the G of rs7200786 are recessive risk alleles. However, these two alleles are protective against multiple sclerosis and rheumatoid arthritis (Vitale et al., 2002; International Multiple Sclerosis Genetics et al., 2007; Martinez et al., 2010; Skinningsrud et al., 2010; International Multiple Sclerosis Genetics et al., 2013). When the relationship between PD and multiple sclerosis remains in debate (Li et al., 2012; Pedemonte et al., 2013; Nielsen et al., 2014), PD risk, indeed, appears to be negatively correlated with rheumatoid arthritis (Bacelis et al., 2021; Li et al., 2021). These results suggest that the risk variants of *CLEC16A* may have differential actual impact on PD and certain autoimmune diseases. Interestingly, a recent GWAS pooling study of European ancestry identified 17 shared susceptibility loci between autoimmune diseases and PD (Witoelar et al., 2017). Similar to our case, these loci have not been previously reported by PD GWAS studies (Obeso et al., 2017), but partially reported by individual polymorphism studies. Thus, a variety in research designs and ethnicities may still be valuable in searching for new genetic risks. Conversely, these findings may need further validation in additional populations. Being noted, the above European study identified none of the *CLEC16A* variants (Witoelar et al., 2017). To understand the discordance, we examined allele frequencies of the five variants in Europeans and East Asians in the gnomAD (https://gnomad.broadinstitute.org/). Interestingly, the two PD-associated *CLEC16A* variants happen to be with inverted major and minor allele distributions in these two ethnicities. In detail, the G allele frequency of rs6498169 is at 0.358 and 0.554, and the A allele frequency of rs7200786 is at 0.461 and 0.660, respectively, in Europeans and East Asians (**Supplementary Table S4**). In this case, the pleiotropic discordance may partially attribute to the ethnicity-associated evolutional divergency in *CLEC16A* genetic variation.

It is known that haplotypes are more powerful for the detection of susceptibility alleles than individual variants (Gabriel et al., 2002). By analyzing the five variants, we identified two haplotypes, AAGAG and AACGG (in the order of rs6498169, rs12708716, rs12917716, rs7200786, and rs2903692), serving as risk factor. Further analysis of the effect polymorphisms suggests that the AAG of rs6498169, rs12708716 and rs2903692 represents the core haplotype associated with PD. Results of the functional annotation analyses suggest that rs6498169 is highly probable to affect transcription factor binding and target gene expression, such as through the modulation of GATA2 and HES1 binding. In contrast, rs7200786 is with relatively less probability to be functional. These results appear to be in line with the above highest-risk haplotype analysis, which shows that the rs6498169 locus, but not rs7200786, is within the core in association with PD. Based on the quantitative trait locus analysis, the *CLEC16A* expression may be affected by rs6498169 in a splicing-regulating way and by rs7200786 in both expression- and splicing-regulating ways. In addition, rs6498169 and rs7200786 may also affect the *RMI2* expression in an expression-regulating way. RMI2 is a

component of the BLM (Bloom syndrome RecQ like helicase) complex and is essential for genome stability (Hudson et al., 2016). However, the effects of these two variants on gene expression appear mainly in brain regions other than the substantia nigra. Indeed, although the substantia nigra is most profoundly affected in PD, other brain regions are also important for PD pathogenesis and clinical manifestations, such as putamen, cortex and amygdala (Kalia and Lang, 2015).

In conclusion, the present study demonstrates that the autoimmune disease associated *CLEC16A* genetic variants are associated with PD susceptibility in Han Chinese. Specifically, the A of rs6498169 and the G of rs7200786 serve as recessive risk alleles towards PD. These findings provide genetic insights into the pleiotropic role of *CLEC16A* and strengthen the link between PD and autoimmune diseases.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The name of the repository and link can be found below: Dryad; https://datadryad.org/stash/share/_yc2e6e8jRgT3dCBI-WN2w4NZNY1CqRkqiSnD4q1h7w.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Second Affiliated Hospital and Yuying Children's Hospital, Wenzhou Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

H-HF and LC designed the research; H-HF, LC, X-XJ, S-SL, K-YW, H-JD, and MM conducted experiments; LC, X-XJ, Y-DS, and H-MW analyzed data; XZ contributed samples; BO helped data analysis and edited the manuscript; LC and J-HZ wrote the manuscript; H-MW, J-HZ, and XZ supervised the study. All the authors have read and approved the final manuscript.

FUNDING

This work was supported by the funding from Zhejiang Provincial Natural Science Foundation (LD19H090001 and LZ19H090002), National Natural Science Foundation of China (81771380, 81771510, and 82071585), Zhejiang Provincial Medical Technology Program (2019KY445), and Wenzhou Medical University (89218014).

ACKNOWLEDGMENTS

The authors are thankful to their colleagues for clinical data collection and management of patients. The authors appreciate all the subjects participating in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.856493/full#supplementary-material

REFERENCES

Bacelis, J., Compagno, M., George, S., Pospisilik, J. A., Brundin, P., Naluai, Å. T., et al. (2021). Decreased Risk of Parkinson's Disease after Rheumatoid Arthritis Diagnosis: A Nested Case-Control Study with Matched Cases and Controls. *Jpd* 11 (2), 821–832. doi:10.3233/JPD-202418

Bonam, S. R., and Muller, S. (2020). Parkinson's Disease Is an Autoimmune Disease: A Reappraisal. *Autoimmun. Rev.* 19 (12), 102684. doi:10.1016/j.autrev.2020.102684

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

Francis, P. J., Schultz, D. W., Hamon, S., Ott, J., Weleber, R. G., and Klein, M. L. (2007). Haplotypes in the Complement Factor H (CFH) Gene: Associations with Drusen and Advanced Age-Related Macular Degeneration. *PLoS One* 2 (11), e1197. doi:10.1371/journal.pone.0001197

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., et al. (2002). The Structure of Haplotype Blocks in the Human Genome. *Science* 296 (5576), 2225–2229. doi:10.1126/science.1069424

Hakonarson, H., Grant, S. F. A., Bradfield, J. P., Marchand, L., Kim, C. E., Glessner, J. T., et al. (2007). A Genome-wide Association Study Identifies KIAA0350 as a Type 1 Diabetes Gene. *Nature* 448 (7153), 591–594. doi:10.1038/nature06010

Hudson, D. F., Amor, D. J., Boys, A., Butler, K., Williams, L., Zhang, T., et al. (2016). Loss of RMI2 Increases Genome Instability and Causes a Bloom-like Syndrome. *Plos Genet.* 12 (12), e1006483. doi:10.1371/journal.pgen.1006483

Hughes, A. J., Daniel, S. E., Kilford, L., and Lees, A. J. (1992). Accuracy of Clinical Diagnosis of Idiopathic Parkinson's Disease: a Clinico-Pathological Study of 100 Cases. *J. Neurol. Neurosurg. Psychiatry* 55 (3), 181–184. doi:10.1136/jnnp.55.3.181

International Multiple Sclerosis Genetics, C., Beecham, A. H., Patsopoulos, N. A., Xifara, D. K., Davis, M. F., Kemppinen, A., et al. (2013). Analysis of Immune-Related Loci Identifies 48 New Susceptibility Variants for Multiple Sclerosis. *Nat. Genet.* 45 (11), 1353–1360. doi:10.1038/ng.2770

International Multiple Sclerosis Genetics, C., Hafler, D. A., Compston, A., Sawcer, S., Lander, E. S., Daly, M. J., et al. (2007). Risk Alleles for Multiple Sclerosis Identified by a Genomewide Study. *N. Engl. J. Med.* 357 (9), 851–862. doi:10.1056/NEJMoa073493

Kalia, L. V., and Lang, A. E. (2015). Parkinson's Disease. *The Lancet* 386 (9996), 896–912. doi:10.1016/S0140-6736(14)61393-3

Li, C., Ou, R., and Shang, H. (2021). Rheumatoid Arthritis Decreases Risk for Parkinson's Disease: a Mendelian Randomization Study. *Npj Parkinsons Dis.* 7 (1), 17. doi:10.1038/s41531-021-00166-x

Li, X., Sundquist, J., and Sundquist, K. (2012). Subsequent Risks of Parkinson Disease in Patients with Autoimmune and Related Disorders: a Nationwide Epidemiological Study from Sweden. *Neurodegener. Dis.* 10 (1-4), 277–284. doi:10.1159/000333222

Lindestam Arlehamn, C. S., Dhanwani, R., Pham, J., Kuan, R., Frazier, A., Rezende Dutra, J., et al. (2020). α-Synuclein-specific T Cell Reactivity Is Associated with Preclinical and Early Parkinson's Disease. *Nat. Commun.* 11 (1), 1875. doi:10.1038/s41467-020-15626-w

Luciano, M., Hagenaars, S. P., Davies, G., Hill, W. D., Clarke, T.-K., Shirali, M., et al. (2018). Association Analysis in over 329,000 Individuals Identifies 116 Independent Variants Influencing Neuroticism. *Nat. Genet.* 50 (1), 6–11. doi:10.1038/s41588-017-0013-8

Márquez, A., Varadé, J., Robledo, G., Martínez, A., Mendoza, J. L., Taxonera, C., et al. (2009). Specific Association of a CLEC16A/KIAA0350 Polymorphism

with NOD2/CARD15– Crohn's Disease Patients. *Eur. J. Hum. Genet.* 17 (10), 1304–1308. doi:10.1038/ejhg.2009.50

Martínez, A., Perdigones, N., Cénit, M. C., Espino, L., Varadé, J., Lamas, J. R., et al. (2010). Chromosomal Region 16p13: Further Evidence of Increased Predisposition to Immune Diseases. *Ann. Rheum. Dis.* 69 (1), 309–311. doi:10.1136/ard.2008.098376

Michel, P. P., Hirsch, E. C., and Hunot, S. (2016). Understanding Dopaminergic Cell Death Pathways in Parkinson Disease. *Neuron* 90 (4), 675–691. doi:10.1016/j.neuron.2016.03.038

Nielsen, N. M., Pasternak, B., Stenager, E., Koch-Henriksen, N., and Frisch, M. (2014). Multiple Sclerosis and Risk of Parkinson's Disease: a Danish Nationwide Cohort Study. *Eur. J. Neurol.* 21 (1), 107–111. doi:10.1111/ene.12255

Nischwitz, S., Cepok, S., Kroner, A., Wolf, C., Knop, M., Müller-Sarnowski, F., et al. (2011). More CLEC16A Gene Variants Associated with Multiple Sclerosis. *Acta Neurol. Scand.* 123 (6), 400–406. doi:10.1111/j.1600-0404.2010.01421.x

Obeso, J. A., Stamelou, M., Goetz, C. G., Poewe, W., Lang, A. E., Weintraub, D., et al. (2017). Past, Present, and Future of Parkinson's Disease: A Special Essay on the 200th Anniversary of the Shaking Palsy. *Mov Disord.* 32 (9), 1264–1310. doi:10.1002/mds.27115

Pedemonte, E., Trabucco, E., Cella, M., and Solaro, C. (2013). Parkinsonism in Multiple Sclerosis Patients: a Casual or Causal Association? *Parkinsonism Relat. Disord.* 19 (4), 492–493. doi:10.1016/j.parkreldis.2012.11.021

Redmann, V., Lamb, C. A., Hwang, S., Orchard, R. C., Kim, S., Razi, M., et al. (2016). Clec16a Is Critical for Autolysosome Function and Purkinje Cell Survival. *Sci. Rep.* 6, 23326. doi:10.1038/srep23326

Schuster, C., Gerold, K. D., Schober, K., Probst, L., Boerner, K., Kim, M.-J., et al. (2015). The Autoimmunity-Associated Gene CLEC16A Modulates Thymic Epithelial Cell Autophagy and Alters T Cell Selection. *Immunity* 42 (5), 942–952. doi:10.1016/j.immuni.2015.04.011

Scott, K. M., Kouli, A., Yeoh, S. L., Clatworthy, M. R., and Williams-Gray, C. H. (2018). A Systematic Review and Meta-Analysis of Alpha Synuclein Auto-Antibodies in Parkinson's Disease. *Front. Neurol.* 9, 815. doi:10.3389/fneur.2018.00815

Skinningsrud, B., Lie, B. A., Husebye, E. S., Kvien, T. K., Forre, O., Flato, B., et al. (2010). A CLEC16A Variant Confers Risk for Juvenile Idiopathic Arthritis and Anti-cyclic Citrullinated Peptide Antibody Negative Rheumatoid Arthritis. *Ann. Rheum. Dis.* 69 (8), 1471–1474. doi:10.1136/ard.2009.114934

Sole, X., Guino, E., Valls, J., Iniesta, R., and Moreno, V. (2006). SNPStats: a Web Tool for the Analysis of Association Studies. *Bioinformatics* 22 (15), 1928–1929. doi:10.1093/bioinformatics/btl268

Soleimanpour, S. A., Gupta, A., Bakay, M., Ferrari, A. M., Groff, D. N., Fadista, J., et al. (2014). The Diabetes Susceptibility Gene Clec16a Regulates Mitophagy. *Cell* 157 (7), 1577–1590. doi:10.1016/j.cell.2014.05.016

Strafella, C., Caputo, V., Termine, A., Assogna, F., Pellicano, C., Pontieri, F. E., et al. (2021). Immune System and Neuroinflammation in Idiopathic Parkinson's Disease: Association Analysis of Genetic Variants and miRNAs Interactions. *Front. Genet.* 12, 651971. doi:10.3389/fgene.2021.651971

Sulzer, D., Alcalay, R. N., Garretti, F., Cote, L., Kanter, E., Agin-Liebes, J., et al. (2017). T Cells from Patients with Parkinson's Disease Recognize α-synuclein Peptides. *Nature* 546 (7660), 656–661. doi:10.1038/nature22815

van de Warrenburg, B. P. C., Church, A. J., Martino, D., Candler, P. M., Bhatia, K. P., Giovannoni, G., et al. (2008). Antineuronal Antibodies in Parkinson's Disease. *Mov. Disord.* 23 (7), 958–963. doi:10.1002/mds.21929

van Luijn, M. M., Kreft, K. L., Jongsma, M. L., Mes, S. W., Wierenga-Wolf, A. F., van Meurs, M., et al. (2015). Multiple Sclerosis-Associated CLEC16A Controls HLA Class II Expression via Late Endosome Biogenesis. *Brain* 138 (Pt 6), 1531–1547. doi:10.1093/brain/awv080

Vitale, E., Cook, S., Sun, R., Specchia, C., Subramanian, K., Rocchi, M., et al. (2002). Linkage Analysis Conditional on HLA Status in a Large North American Pedigree Supports the Presence of a Multiple Sclerosis Susceptibility Locus on Chromosome 12p12. *Hum. Mol. Genet.* 11 (3), 295–300. doi:10.1093/hmg/11.3.295

Ward, L. D., and Kellis, M. (2016). HaploReg V4: Systematic Mining of Putative Causal Variants, Cell Types, Regulators and Target Genes for Human Complex Traits and Disease. *Nucleic Acids Res.* 44 (D1), D877–D881. doi:10.1093/nar/gkv1340

Witoelar, A., Jansen, I. E., Wang, Y., Desikan, R. S., Gibbs, J. R., Blauwendraat, C., et al. (2017). Genome-wide Pleiotropy between Parkinson Disease and Autoimmune Diseases. *JAMA Neurol.* 74 (7), 780–792. doi:10.1001/jamaneurol.2017.0469

Zappia, M., Crescibene, L., Bosco, D., Arabia, G., Nicoletti, G., Bagalà, A., et al. (2002). Anti-GM1 Ganglioside Antibodies in Parkinson's Disease. *Acta Neurol. Scand.* 106 (1), 54–57. doi:10.1034/j.1600-0404.2002.01240.x

Zou, M., Li, R., Wang, J.-Y., Wang, K., Wang, Y.-N., Li, Y., et al. (2018). Association Analyses of Variants of SIPA1L2, MIR4697, GCH1, VPS13C, and DDRGK1 with Parkinson's Disease in East Asians. *Neurobiol. Aging* 68, e7–159. doi:10.1016/j.neurobiolaging.2018.03.005

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

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

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 08 June 2022
doi: 10.3389/fgene.2022.913196

Contribution of HLA and KIR Alleles to Systemic Sclerosis Susceptibility and Immunological and Clinical Disease Subtypes

Aimee L. Hanson[1,2], Joanne Sahhar[3], Gene-Siew Ngian[3], Janet Roddy[4], Jennifer Walker[5], Wendy Stevens[6,7], Mandana Nikpour[6,7], Shervin Assassi[8], Susanna Proudman[9,10], Maureen D. Mayes[8], Tony J. Kenna[11†] and Matthew A. Brown[12,13†]*

[1]Department of Medicine, University of Queensland, Brisbane, QLD, Australia, [2]Department of Medicine, University of Cambridge, Cambridge, United Kingdom, [3]Department of Medicine, Clayton and Monash Health, Monash University, Melbourne, VIC, Australia, [4]Department of Rheumatology, Royal Perth Hospital, Perth, WA, Australia, [5]Rheumatology Unit, Flinders Medical Centre, Adelaide, SA, Australia, [6]Department of Medicine, University of Melbourne, St Vincent's Hospital, Melbourne, VIC, Australia, [7]Department of Rheumatology, St Vincent's Hospital, Melbourne, VIC, Australia, [8]Division of Rheumatology, University of Texas, Houston, TX, United States, [9]Rheumtology Unit, Royal Adelaide Hospital, Adelaide, SA, Australia, [10]Discipline of Medicine, University of Adelaide, Adelaide, SA, Australia, [11]Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, QLD, Australia, [12]Genomics England, Charterhouse Square, London, United Kingdom, [13]Department of Medical and Molecular Genetics, Faculty of Life Sciences and Medicine, King's College London, London, United Kingdom

OPEN ACCESS

Edited by:
Bassam R. Ali,
United Arab Emirates University,
United Arab Emirates

Reviewed by:
Jukka Partanen,
Finnish Red Cross Blood Service, Finland
Masao Ota,
Shinshu University, Japan

***Correspondence:**
Matthew A. Brown
matt.brown@kcl.ac.uk

[†]These authors have contributed equally to this work and share senior authorship

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

Received: *05 April 2022*
Accepted: *03 May 2022*
Published: *08 June 2022*

Citation:
Hanson AL, Sahhar J, Ngian G-S, Roddy J, Walker J, Stevens W, Nikpour M, Assassi S, Proudman S, Mayes MD, Kenna TJ and Brown MA (2022) Contribution of HLA and KIR Alleles to Systemic Sclerosis Susceptibility and Immunological and Clinical Disease Subtypes. Front. Genet. 13:913196. doi: 10.3389/fgene.2022.913196

Systemic sclerosis (SSc) is an autoinflammatory, fibrotic condition of unknown aetiology. The presence of detectable autoantibodies against diverse nuclear antigens, as well as strong HLA associations with disease, suggest autoimmune involvement, however the links between endogenous and exogenous risk factors and SSc pathology remain undetermined. We have conducted a genetic analysis of *HLA* inheritance in two independent and meta-analysed cohorts of 1,465 SSc cases and 13,273 controls, including stratified association analyses in clinical and autoantibody positive subgroups of disease. Additionally, we have used patient genotypes to impute gene dosages across the *KIR* locus, encoding paired activating and inhibitory lymphocyte receptors for Class I HLA ligands, to conduct the largest analysis of *KIR-HLA* epistatic interactions in SSc to date. We confirm previous Class II HLA associations with SSc risk and report a new Class I association with haplotype *HLA-B*44:03-HLA-C*16:01* at genome-wide significance (GWS). We further report statistically significant *HLA* associations with clinical and serological subtypes of disease through direct case-case comparison, and report a new association of *HLA-DRB1*15:01*, previously shown to bind topoisomerase-1 derived peptides, with anti-topoisomerase (ATA) positive disease. Finally, we identify genetic epistasis between KIRs and HLA class I ligands, suggesting genetic modulation of lymphocyte activation may further contribute to an individual's underlying disease risk. Taken together, these findings support future functional investigation into endogenous immunological and environmental stimuli for disrupted immune tolerance in SSc.

Keywords: systemic sclerosis (scleroderma), HLA association and disease, human disease genetics, killer immunoglobulin like receptor (KIR), immunogenetics

INTRODUCTION

Systemic sclerosis (SSc; also known as scleroderma), is a connective-tissue disease of heterogeneous clinical presentation, characterised by a complex interplay between autoinflammatory and autoimmune processes, tissue fibrosis and vascular injury. Pathophysiology associated with SSc has been attributed to uncontrolled inflammation and associated activation of tissue resident fibroblasts, excessive collagen production and extracellular matrix remodelling (LeRoy, 1974). Two major subsets of disease are differentiated by the extent of skin and organ involvement. Diffuse cutaneous SSc (dcSSc) is characterised by extensive skin fibrosis and increased mortality associated with severe internal organ pathology. Alternatively, in limited cutaneous disease (lcSSc), skin involvement is typically restricted to the face and distal extremities, and vascular complications predominate. Investigated triggers of the inflammatory, fibrotic cascade in SSc suggest a role for a provocative environmental exposure, ranging from occupational chemicals (Sluis-Cremer et al., 1985; Kettaneh et al., 2007) to microbial infection (Maul et al., 1989; Lunardi et al., 2006), acting upon a background of heightened genetic predisposition. Albeit, the precise series of exogenous and endogenous circumstances that culminate in disease, and primary differentiators of clinical progression, remain unknown.

Anti-nuclear antibodies (ANA) targeting a range of cellular proteins are detected in the sera of up to 95% of SSc patients (Beck et al., 1963; Mierau et al., 2011), some of which have diagnostic utility in distinguishing SSc from other connective tissue diseases. Of the most common, anti-centromere antibodies (ACA), seen in 20–30% of all SSc cases, demonstrate reactivity against centromeric nucleoproteins such as the DNA-binding protein Centromere Protein B (CENP-B), and are predictive of limited disease with lower frequency of pulmonary fibrosis and associated mortality (Steen et al., 1988; Ho and Reveille, 2003). Anti-Scl70 antibodies directed against an epitope of the topoisomerase enzyme (abbreviated anti-topoisomerase, ATA; seen in 15–20% of SSc) are conversely predictive of diffuse disease and found in ~45% of those who progress with lung complications (Reveille et al., 2003). Additionally, ATAs are associated with severity and development of interstitial lung disease (ILD), the leading cause of SSc mortality, in both limited and diffuse disease (Assassi et al., 2010; Jandali et al., 2022). A third and heterogeneous group of anti-nucleolar antibodies (ANoA; reported in 15–40% of patients) target, among other autoantigens, exosome, and ribonuclease components and several RNA polymerases (RNAP), with anti-RNAP antibodies demonstrating a strong association with heart failure and increased mortality in dcSSc (Jacobsen et al., 2001). Intriguingly, the co-occurrence of multiple ANAs of unique specificity in a single patient is rare. Their near mutual exclusivity suggests a unique interplay between genetic and environmental factors in the onset and progression of distinct clinical subclasses of disease, though whether ANA-associated autoimmunity is a cause or consequence of pathological processes in SSc remains undetermined.

The association of allelic variation across the class II human leukocyte antigen (HLA) locus with SSc risk is well established, supporting a functional role for CD4 T cell mediated autoantibody production in response to HLA class II restricted autoantigens in genetically susceptible individuals. Different HLA alleles demonstrate disease association when tested in lcSSc or dcSSc cohorts, or those stratified by autoantibody profile, implying that genetic heterogeneity significantly underpins the clinical heterogeneity of this disease. The strong risk association of *HLA-DRB1*11:04* and *HLA-DPB1*13:01* with SSc has been reported in multiple ethnic groups, enriched specifically in ATA + patients (Fanning et al., 1998; Arnett et al., 2010; Gourh et al., 2020; Acosta-Herrera et al., 2021). Conversely, *HLA-DQA1*01:01/4* and *HLA-DQB1*05* have been associated with risk of disease in ACA + Caucasians and Hispanics (Fanning et al., 1998; Reveille et al., 2001), and protective associations with *HLA-DRB1*07:01* in this serological subgroup are described in multiple European cohorts (Zochling et al., 2014; Gourh et al., 2020; Acosta-Herrera et al., 2021).

Although less frequently reported, HLA class I associations with SSc suggest an additional role for $CD8^+$ T cell mediated autoimmunity in disease. Early studies show enrichment of the *HLA-A*09* subgroup allele *HLA-A*24* in dcSSc, and *HLA-A*30* and *HLA-A*32* risk associations have also been reported in Caucasian and Brazilian populations, the former associated with risk of pulmonary hypertension and pulmonary fibrosis in SSc patients (Gladman et al., 2005; Del Rio et al., 2013). A recent intensive analysis of 9,095 SSc patients identified a significant disease risk association with *HLA-B*08:01*, independent of those observed with HLA Class II alleles (Acosta-Herrera et al., 2021). Increased $CD8^+$ T cell clonality observed in the lungs and blood of SSc patients suggests proliferation in response to an immunogenic antigen (Yurovsky et al., 1996; Servaas et al., 2021), and activated $CD8^+$ T cells have been identified in the skin, and isolated from the fibrotic lungs of SSc patients, where they express pro-fibrotic cytokines which stimulate fibroblast proliferation and collagen production (Atamas et al., 1999; Li et al., 2017). A summary of the findings from large-scale cohorts reporting classical HLA associations with SSc and disease subgroups can be found in **Supplementary Table S1**.

Additional to their role in antigen presentation, Class I HLA ligands are recognised by killer immunoglobulin-like receptors (KIRs) expressed on NK and $CD8^+$ T cells, with roles in buffering lymphocyte activation and safeguarding against innate NK killing activity. Copy number variable haplotypes containing diverse combinations of 15 unique KIR genes encode activating (KIR2DS and KIR3DS) and inhibitory (KIR2DL and KIR3DL) receptors that engage specific subtypes of HLA Class I alleles. Canonically, KIR3DL1 is an inhibitory receptor for HLA-Bw4 ligands, a subgroup of HLA-A and B alleles carrying a recognised amino acid motif at positions 77–83 in the *a*-helix (Gumperz et al., 1996). The strength of this inhibitory interaction is governed by position 80 of the HLA Class I allele, being strongest for those alleles carrying isoleucine (Ile80), and weaker for those carrying threonine (The80), at position 80. Conversely, HLA-Bw6 alleles, lacking the Bw4 motif, are not known to serve as KIR ligands.

KIR2DL2 and 2DL3 recognise the HLA-C1 family of alleles characterised by a position 80 asparagine, and KIR2DL1 engages members of the HLA-C2 allelic family, which carry a position 80 lysine (Winter et al., 1998) (**Supplementary Table S2** lists major HLA-C, -Bw4 and -Bw6 family members). Ligands that engage activating KIRs are less well defined and extend beyond class I HLA to other surface markers of cell stress. However, HLA-dependent, peptide-specific activation of NK cells through KIR2DS1 recognition of HLA-C2 has been demonstrated *in vitro* (Chewning et al., 2007), and KIR3DS1 has been shown to engage HLA-B*57 (HLA-Bw4 subclass) bound HIV-derived peptides to activate NK killing responses (O'Connor et al., 2015). The exacerbated diversity attributed to copy number and allelic hyper-polymorphism across both the *KIR* genes and the combination of HLA ligands inherited imparts extensive variability in immunologic potential across the human population. Accordingly, *KIR-HLA* associations have been reported to play a role in genetic risk for many immune-mediated and infectious diseases (Khakoo et al., 2004; Pellett et al., 2007; Fusco et al., 2010; Hanson et al., 2020). Increased frequency of activating receptor *KIR2DS1* in SSc patients, particularly in those who lack a canonical HLA-C2 ligand for the paired inhibitory receptor *KIR2DL1,* has been reported (Pellett et al., 2007). Further studies report that lack of *KIR2DL2* inheritance increases SSc risk (Salim et al., 2010), and that coinheritance of canonical inhibitory receptor-ligand pair *KIR3DL1-HLA-Bw4*(The80) protects from disease (Mahmoudi et al., 2017), suggesting that activating KIR inheritance in the absence of compatible inhibitory interacting pairs may result in poorly restrained NK and T-cell activation and associated damaging inflammation.

Here we present the findings of an extensive investigation into *HLA* and *KIR* inheritance in two independent cohorts of SSc patients, including meta-analysis of 1,465 SSc cases and 13,273 controls. We present stratified *HLA*-association analyses in clinical and autoantibody positive SSc subgroups to dissect the genetic contribution of HLA alleles to unique serological and clinical manifestations of disease. We further impute *KIR* gene content information from participant genotypes to conduct the largest analysis of *KIR-HLA* coinheritance in this disease to date, and the first to address epistatic KIR interactions with both HLA subtypes and alleles in SSc. It is hoped that more thorough delineation of the genetic contributors to SSc will orient future research into the precise underlying triggers and immunological mechanisms of disease, and inform the development of targeted treatments to reduce the associated morbidity and mortality of those diagnosed.

MATERIALS AND METHODS

Study Cohorts

Two independent cohorts of genotyped SSc patients and ethnically matched controls were used to test for genetic associations with disease and clinical disease subgroups across the *HLA* and *KIR* loci. Following quality filtering as described below, Cohort 1 comprised 503 Australian SSc patients recruited by the Australian Scleroderma Interest Group (ASIG), and Cohort 2 comprised 962 SSc patients from the United States of America, a subset of a previously published discovery cohort (Mayes et al., 2014). A total of 13,858 healthy controls, originally recruited by the International Genetics of Ankylosing Spondylitis (IGAS) Consortium as previously reported (Cortes et al., 2013), were split evenly between both SSc cohorts, with 6,632 and 6,641 controls remaining in Cohorts 1 and 2 respectively following quality filtering. Meta-analysis was performed in a combined cohort of 1,465 SSc cases and 13,273 controls. All patients met the American College of Rheumatology criteria for clinical diagnosis of SSc (Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee, 1980), or the Medsger criteria for limited SSc (LeRoy and Medsger, 2001). Clinical metadata, including limited (lcSSc) or diffuse (dcSSc) disease classification, and ACA, ATA and ANoA autoantibody status (present/absent) was available for 490 Cohort 1 patients, and all 962 of Cohort 2 (classification frequencies shown in **Supplementary Figure S1**). Written informed consent was obtained from all participants, with research ethics approval granted by the relevant ethics committee at each participating centre.

Genotyping and Sample Filtering

All SSc cases and controls were genotyped on the Illumina Immunochip array, which has high density SNP coverage of the leukocyte receptor complex at chromosome 19q13.4, where the *KIR* and *LILR* genes are encoded (Cortes and Brown, 2011). Sample genotypes were merged with genotypes from HapMap reference human populations and 17,374 common autosomal SNPs outside regions of long-range linkage disequilibrium (LD) were used to conduct principal component analysis (PCA) with shellfish (https://www.stats.ox.ac.uk/~davison/software/shellfish/shellfish.php) for ethnicity assessment. Only patients and controls falling within plus or minus 5 standard deviations from the mean of the European sample cluster were retained. Principal components (PC) were recalculated for the filtered European participants and the first ten PCs fitted as covariates in all regression models to correct for remaining population stratification.

HLA Imputation

Imputation of 268 classical HLA alleles to four-digit resolution was performed using HLA*IMP:03 (http://imp.science.unimelb.edu.au/hla/) (Motyer et al., 2016). The estimated imputation accuracy of Class I and II loci ranged between 95% (HLA-DRB1) and 99.84% (HLA-DPA1), with all alleles with a minor allele frequency >1% imputed with a mean estimated accuracy of 95.5% (IQR 94.7–98.8%). Consequently, no posterior probability threshold was applied to imputed HLA allele calls. The HLA subclass of each Class I allele (HLA-Bw4, Bw6, C1 or C2) was assigned based on known allele groupings (**Supplementary Table S2**) and the number of alleles carrying HLA-Bw4 (total), HLA-Bw4(I80), HLA-Bw4(T80), HLA-C1 or HLA-C2 motifs were summed for each individual for use in statistical analysis.

KIR Imputation

A total of 265 and 224 SNPs spanning the *KIR* locus (Chr19: 59,793,991–60,190,556, Hg18) were available for SSc patients in Cohort 1 and Cohort 2 respectively, owing to independent genotype quality filtering prior to data acquisition. Matching variant positions were extracted from the paired controls of each cohort to minimise bias in *KIR* imputation accuracy, and SNP minor allele frequency was confirmed to closely match that of the KIR*IMP reference cohort (Vukcevic et al., 2015). Imputation of gene dosages across 14 copy-number variable *KIR* genes (*KIR2DP1, 2DS1, 2DS2, 2DS3, 2DS4, 2DS5, 2DL1, 2DL2, 2DL3, 2DL4, 2DL5, 3DP1, 3DL1, 3DS1*) was performed separately for Cohort 1 and 2 by passing phased SNP haplotypes to KIR*IMP (http://imp.science.unimelb.edu.au/kir/) (Vukcevic et al., 2015). Isoforms of *KIR3DL1*, with variable inclusion of exons 4 or 9, were distinguished by the imputation algorithm, but only total *KIR3DL1* gene dosage was used in analyses throughout. Presence or absence of a 22bp deletion in the *KIR2DS4* gene was denoted by *KIR2DS4DEL* and *KIR2DS4WT* dosages respectively, with *KIR2DS4TOTAL* capturing summed gene dosage. The posterior probability of *KIR* imputation accuracy was comparable across patients and controls from both study cohorts (**Supplementary Figure S2**). *KIR* genes present on rarer and more copy-number variable haplotypes (*KIR2DP1, 2DL1, 2DL5,* and *2DS3*) showed reduced imputation accuracy (**Supplementary Figure S2**), and consequently individuals predicted to carry rare haplotypes exhibited reduced haplotype posterior probability scores (as designated by the *KIRhaplotype* metric returned from the KIR*IMP software). Statistical analyses were conducted on both the full complement of imputed haplotypes, and those imputed above range of *KIRhaplotype* thresholds (0.5–0.9). Imputation thresholding skewed represented haplotypes toward those more common in the population (and more accurately imputed) but did not have a considerable effect on the results of *KIR-HLA* interaction analyses, which are gene not haplotype based, save for a reduction in statistical power. Thus, no posterior probability cut-off was applied to imputed haplotypes in the analyses presented here. To validate imputed haplotype calls on the population level, imputed *KIR* gene frequencies and percent prevalence across Cohorts 1 and 2 were compared to those reported for European populations in the Allele Frequency Net Database (Gonzalez-Galarza et al., 2020) (**Supplementary Figure S3**). *KIR* gene content haplotypes were annotated in accordance with the KIR*IMP reference cohort and imputed haplotype frequencies were compared to reference cohort frequencies derived from 793 nuclear families from the US and United Kingdom (Jiang et al., 2012) (**Supplementary Figure S4**).

Statistical Analyses

Genotype associations with disease (or disease subgroup) status were assessed using logistic regression with the glm()function in R (R Core Team, 2021) for both independent and meta-analysis cohorts. The first 10 principal components capturing population genetic diversity (as detailed above) were included as covariates in each model. The association of disease status with HLA alleles at 4-digit resolution was assessed under a dominant inheritance model (allele absent = 0, heterozygote or homozygote = 1), and *KIR* gene dosage under both a dominant and recessive inheritance model (allele absent or heterozygote = 0, homozygote = 1), eg:

$$\begin{aligned}&\mathrm{glm}(\mathrm{status} \sim \mathrm{allele.count} + \mathrm{PC1} + \mathrm{PC2} + \mathrm{PC3} + \mathrm{PC4} + \mathrm{PC5}\\&\quad + \mathrm{PC6} + \mathrm{PC7} + \mathrm{PC8} + \mathrm{PC9} + \mathrm{PC10}, \mathrm{family}\\&\quad = \mathrm{binomial}(\mathrm{link} = \mathrm{logit}))\end{aligned}$$

Iterative conditional analysis was used to dissect independent HLA associations across the locus exhibiting strong LD, in the combined meta-analysis cohort only. Here, the most significant disease-associated allele was added as a covariate to the logistic regression and association analyses repeated sequentially until no HLA allele showed a disease association below the GWS threshold $p<5 \times 10^{-8}$. Pairwise conditional analysis was conducted by conditioning each disease associated allele upon every other in a pairwise fashion, eg:

$$\begin{aligned}&\mathrm{glm}(\mathrm{status} \sim \mathrm{allele.count.1} + \mathrm{allele.count.2} + \mathrm{PC1} + \mathrm{PC2} + \mathrm{PC3}\\&\quad + \mathrm{PC4} + \mathrm{PC5} + \mathrm{PC6} + \mathrm{PC7} + \mathrm{PC8} + \mathrm{PC9} + \mathrm{PC10}, \mathrm{family}\\&\quad = \mathrm{binomial}(\mathrm{link} = \mathrm{logit}))\end{aligned}$$

HLA associations were also assessed in clinical subgroups of lcSSc and dcSSc patients, both by comparing allele frequencies in each to healthy controls, and between disease states, using logistic regression as detailed above. HLA allele frequencies in ANA positive patients (split into subsets of ACA, ATA and ANoA positive disease) were compared to the corresponding autoantibody negative patient cohort. The interaction between *KIR* gene and HLA subclass carriage in SSc was assessed by inclusion of an interaction term in the logistic regression, treating *KIR* and *HLA* inheritance as either dominant or recessive in every combination (i.e. KIR_{DOM} x HLA_{DOM}, KIR_{DOM} x HLA_{REC}, KIR_{REC} x HLA_{DOM}, KIR_{REC} x HLA_{REC}), eg:

$$\begin{aligned}&\mathrm{glm}(\mathrm{status} \sim \mathrm{HLA.count} + \mathrm{KIR.count} + \mathrm{HLA.count}^{*}\mathrm{KIR.count}\\&\quad + \mathrm{PC1} + \mathrm{PC2} + \mathrm{PC3} + \mathrm{PC4} + \mathrm{PC5} + \mathrm{PC6} + \mathrm{PC7} + \mathrm{PC8} + \mathrm{PC9}\\&\quad + \mathrm{PC10}, \mathrm{family}\\&\quad = \mathrm{binomial}(\mathrm{link} = \mathrm{logit}))\end{aligned}$$

Due to the large number of statistical tests conducted to assess epistatic interaction between these loci, only those KIR-HLA interactions achieving nominal significance (p_{int} <0.05) in the meta-analysis cohort, and occurring between a *KIR* and a disease associated class I HLA ligands, are reported in the main text, with all findings between biologically validated receptor-ligand pairs included as supplementary data as referenced below.

RESULTS

HLA Associations With SSc

Class I and II HLA associations with SSc were assessed using logistic regression, under a dominant inheritance model.

TABLE 1 | HLA associations with SSc

	SSc vs. control													
	Meta-Analysis cohort				Cohort 1				Cohort 2					
HLA allele	SSc prop.(Count)	CO prop.(Count)	OR	*p*	SSc prop.(Count)	CO prop.(Count)	OR	*p*	SSc prop.(Count)	CO prop.(Count)	OR	*p*
*DRB1*11:04*	0.119 (174/1465)	0.035 (459/13273)	2.81	$2.3x10^{-25}$	0.099 (50/503)	0.035 (230/6632)	2.36	$6.4x10^{-7}$	0.129 (124/962)	0.034 (229/6641)	3.08	$1.3x10^{-19}$
*DRB1*07:01*	0.152 (223/1465)	0.257 (3406/13273)	0.52	$1.5 x10^{-17}$	0.157 (79/503)	0.262 (1737/6632)	0.52	$2.5x10^{-7}$	0.150 (144/962)	0.251 (1669/6641)	0.53	$4.4x10^{-11}$
*DQA1*02:01*	0.153 (224/1465)	0.256 (3397/13273)	0.53	$3.5 x10^{-17}$	0.159 (80/503)	0.261 (1733/6632)	0.53	$4.8x10^{-7}$	0.150 (144/962)	0.251 (1664/6641)	0.53	$5.4x10^{-11}$
*DQB1*02:02*	0.102 (149/1465)	0.186 (2468/13273)	0.49	$1.7 x10^{-15}$	0.103 (52/503)	0.193 (1282/6632)	0.47	$5.9x10^{-7}$	0.101 (97/962)	0.179 (1186/6641)	0.51	$2.4x10^{-9}$
*DRB4*01:01*	0.111 (162/1465)	0.188 (2491/13273)	0.54	$9.3 x10^{-13}$	0.109 (55/503)	0.194 (1289/6632)	0.50	$2.6x10^{-6}$	0.111 (107/962)	0.181 (1202/6641)	0.57	$2.9x10^{-7}$
*DPB1*13:01*	0.076 (111/1465)	0.033 (444/13273)	2.20	$1.8 x10^{-12}$	0.068 (34/503)	0.035 (231/6632)	1.85	0.001	0.080 (77/962)	0.032 (213/6641)	2.50	$7.5x10^{-11}$
*B*44:03*	0.050 (73/1465)	0.105 (1397/13273)	0.45	$1.8 x10^{-10}$	0.066 (33/503)	0.110 (728/6632)	0.57	0.002	0.042 (40/962)	0.101 (669/6641)	0.40	$4.9x10^{-8}$
*C*16:01*	0.038 (55/1465)	0.082 (1090/13273)	0.44	$6.2 x10^{-9}$	0.050 (25/503)	0.086 (573/6632)	0.55	0.004	0.031 (30/962)	0.078 (517/6641)	0.39	$9.1x10^{-7}$
*DQA1*05:01*	0.508 (744/1465)	0.421 (5583/13273)	1.34	$1.7 x10^{-7}$	0.509 (256/503)	0.414 (2747/6632)	1.40	0.0003	0.507 (488/962)	0.427 (2836/6641)	1.29	0.0003

CO = control, OR = odds ratio, p = p-value.

Applying a GWS threshold of $p<5 \times 10^{-8}$, significant disease-risk associations were observed with *HLA-DRB1*11:04* (OR = 2.81, $p = 2.3 \times 10^{-25}$) and *HLA-DPB1*13:01* (OR = 2.20, $p = 1.8 \times 10^{-12}$), and protective associations with *HLA-DRB1*07:01* (OR = 0.52, $p = 1.5 \times 10^{-17}$), *HLA-DQA1*02:01* (OR = 0.53, $p = 3.5 \times 10^{-17}$), *HLA-DQB1*02:02* (OR = 0.49, $p = 1.7 \times 10^{-15}$), *HLA-DRB4*01:01* (OR = 0.54, $p = 9.3 \times 10^{-13}$), *HLA-B*44:03* (OR = 0.45, $p = 1.8 \times 10^{-10}$), and *HLA-C*16:01* (OR = 0.44, $p = 6.2 \times 10^{-9}$) in the meta-analysis cohort. A suggestive association was also observed with *HLA-DQA1*05:01* (OR = 1.34, $p = 1.7 \times 10^{-7}$). HLA disease-association signals detected in the meta-analysis cohort were supported in both independent study cohorts 1 and 2, six of which reached GWS in the latter and larger of the two (**Table 1**). The extended table of HLA allele frequencies and associations for all three cohorts is provided in **Supplementary Table S3**).

HLA conditional analysis was performed on the meta-analysis cohort by sequential inclusion of the most significant SSc-associated HLA allele in the logistic regression model. Conditioning on the top disease-risk associated variant, *HLA-DRB1*11:04*, did not ablate significance of any of the remaining disease-associated class I or II alleles, apart from moderate reduction in significance of *HLA-DQA1*05:01* (to $p = 0.004$). Additional correction for the second most strongly associated allele, *HLA-DRB1*07:01*, ablated the protective association with class II alleles *HLA-DQA1*02:01, DQB1*02:02,* and *DRB4*01:01* ($p > 0.04$), and, to a lesser extent, that of class I alleles *HLA-B*44:*03 ($p = 2.0 \times 10^{-4}$) and *HLA-C*16:01* ($p = 3.0 \times 10^{-4}$), suggesting linkage of these loci within a common disease-protective haplotype. The residual disease-association signal seen for risk variant *HLA-DRB1*13:01* remained significant ($p = 1.2 \times 10^{-16}$), and upon correction for this allele no residual HLA association was observed (**Table 2**). Pairwise conditional analysis further demonstrated that either *HLA-DRB1*07:01* or *DQA1*02:01* is the lead associated allele in the protective haplotype also containing *DQB1*02:02* and *DRB4*01:01*, as conditioning on either entirely abolished the association signal with the remaining three alleles. Strong LD was also demonstrated between class I alleles *HLA-B*44:03* and *HLA-C*16:01*; conditioning on the former abolished the *HLA-C* association (**Figure 1**).

HLA Associations With SSc Disease Subgroups

HLA associations were further assessed in subgroups of SSc patients differentiated by autoantibody subclass and limited or diffuse disease subtype (**Table 3**, **Supplementary Tables S4–S9**). *p*-values reported in the text below are derived from the meta-analysis cohort. The disease-risk association of *HLA-DPB1*13:01* was strongest in the diffuse disease cohort (dcSSc OR = 3.2, $p = 2.4 \times 10^{-13}$; lcSSc OR = 1.75, $p = 0.0001$), whereas the protective associations of linked alleles *HLA-DRB1*07:01* and *DQA1*02:01*, *DQB1*02:02* and *DRB4*01:01*, as well as *HLA-B*44:03* and *HLA-C*16:01,* were most pronounced in limited SSc (lcSSc OR<0.45, $p < 2.9 \times 10^{-8}$; dcSSc OR>0.62, $p > 0.01$). The disease risk association with *HLA-DRB1*11:04* remained significant when assessed in both the limited (OR = 2.76, $p_{FDR} = 6.5 \times 10^{-18}$)

TABLE 2 | Conditional HLA association analysis in the meta-analysis cohort.

	DRB1*11:04		DRB1*11:04 + DRB1*07:01		DRB1*11:04 + DRB1*07:01 + DPB1*13:01	
HLA allele	**OR**	***p***	**OR**	***p***	**OR**	***p***
*DRB1*11:04*	—	—	—	—	—	—
*DRB1*07:01*	0.55	$6.1x10^{-15}$	—	—	—	—
*DQA1*02:01*	0.56	$1.4x10^{-14}$	2.70	0.3	2.85	0.2
*DQB1*02:02*	0.52	$3.0x10^{-13}$	0.74	0.04	0.88	0.4
*DRB4*01:01*	0.57	$1.0x10^{-10}$	0.80	0.05	0.84	0.1
*DPB1*13:01*	2.22	$1.4x10^{-12}$	2.60	$1.2x10^{-16}$	—	—
*B*44:03*	0.47	$1.1x10^{-9}$	0.61	0.0002	0.63	0.0004
*C*16:01*	0.46	$3.0x10^{-8}$	0.59	0.0003	0.61	0.0008
*DQA1*05:01*	1.18	$3.8x10^{-3}$	1.09	0.1	1.09	0.1

p-*values below the GWS threshold are underlined. Conditioning alleles are shown at top of table. CO = control, OR = odds ratio, p =* p-*value.*

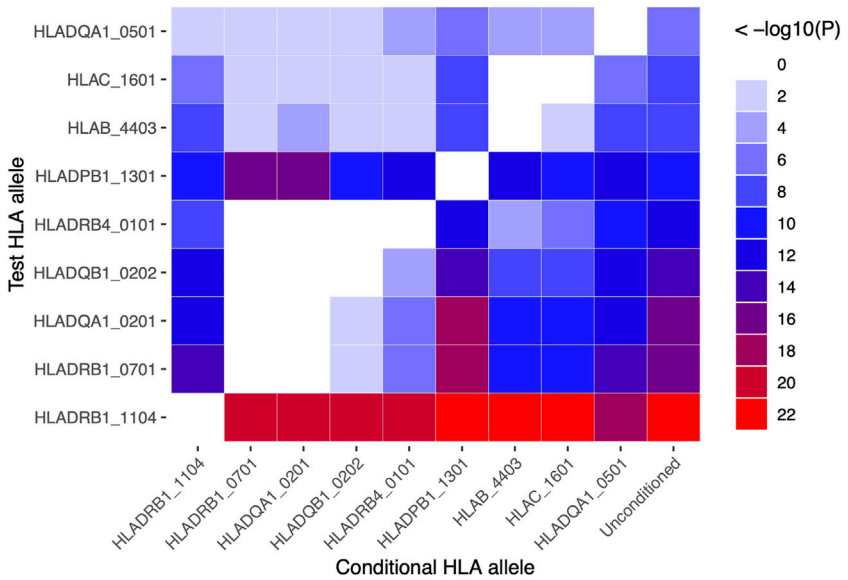


FIGURE 1 | Pairwise conditional analysis of GWS and suggestive HLA class I and II associations in the SSc meta-analysis cohort. The *p*-value for the genetic association of each allele (y-axis) with disease upon correction for each allele (x-axis) in turn is denoted by the colour key on far right. Unconditioned *p*-values are shown in the right column for comparison.

and diffuse disease cohorts (OR = 2.93, p = 3.3×10^{-12}). When allele frequencies were compared directly between lcSSc and dcSSc patients, the disease protective class II alleles *HLA-DRB1*07:01*, *DQA1*02:01* and *DQB1*02:02* were seen at lower frequency in lcSSc than dcSSc, albeit not at GWS with $p < 8.0 \times 10^{-5}$ (**Table 3**). Extended tables of HLA allele associations calculated between disease subsets and controls, and between lcSSc and dcSSc, for all three cohorts are provided in **Supplementary Tables S4–S6**.

Considering associations with autoantibodies, *HLA-DQA1*01:01*, *DQB1*05:01* and *DRB1*01:01* were seen at significantly increased frequency in ACA + compared with ACA − SSc (*HLA-DQA1*01:01* OR = 2.58, $p = 4.5 \times 10^{-15}$; *HLA-DQB1*05:01* OR = 2.39, $p = 3.2 \times 10^{-12}$; *HLA-DRB1*01:01* OR = 2.15, $p = 1.5 \times 10^{-8}$ in the meta-analysis cohort), and reduced frequency in ATA + relative to ATA − disease (*HLA-DQA1*01:01* OR = 0.3, $p_{FDR} = 2.4 \times 10^{-9}$; *HLA-DQB1*05:01* OR = 0.38, $p_{FDR} = 1.7 \times 10^{-6}$; *HLA-DRB1*01:01* OR = 0.42, $p = 0.0001$; **Table 3**). The *HLA-DQA1*01:01* and *DQB1*05:01* alleles were seen at 43.6 and 37.9% respectively in ACA + SSc relative to only 13.3% (both) in ATA + disease and 27 and 23% respectively in controls. *HLA-DQA1*01:01* was also the top HLA to show a suggestive association with ANoA + disease, where it was carried at a reduced frequency relative to ANoA − disease (OR = 0.48, $p = 9.2 \times 10^{-7}$). Alternatively, suggestive disease risk associations were seen with *HLA-DQA1*05:01* and *DQB1*02:01* in the ANoA + cohort (*HLA-DQA1*05:01* OR = 1.84, $p = 1.6 \times 10^{-6}$; *HLA-DQB1*02:01* OR = 1.88, $p = 4.7 \times 10^{-6}$), and the class 1 allele *HLA-C*07:01* was seen at increased frequency relative to ANoA − disease, albeit not at GWS (OR = 1.78, $p = 5.8 \times 10^{-6}$; **Table 3**).

The independent SSc disease-risk association with alleles *HLA-DPB1*13:01* and *HLA-DRB1*11:04* appeared to be driven

TABLE 3 | HLA associations with SSc disease and major autoantibody subgroups.

	Metanalysis						Cohort 1						Cohort 2					
							ATA + vs –											
HLA allele	**ATA– Prop.(Count)**		**ATA+ prop.(Count)**		**OR**	***p***	**ATA– Prop.(Count)**		**ATA+ prop.(Count)**		**OR**	***p***	**ATA– Prop.(Count)**		**ATA+ prop.(Count)**		**OR**	***p***
DRB1*07:01	0.126	(123/974)	0.257	(3406/13273)	0.42	2.2×10^{-18}	0.150	(53/353)	0.262	(1737/6632)	0.49	3.2×10^{-6}	0.113	(70/621)	0.251	(1669/6641)	0.39	3.3×10^{-13}
DQA1*02:01	0.127	(124/974)	0.256	(3397/13273)	0.43	5.3×10^{-18}	0.153	(54/353)	0.261	(1733/6632)	0.51	6.1×10^{-6}	0.113	(70/621)	0.251	(1664/6641)	0.39	3.9×10^{-13}
DRB1*11:04	0.114	(111/974)	0.035	(459/13273)	2.76	6.5×10^{-18}	0.076	(27/353)	0.035	(230/6632)	1.94	0.003	0.135	(84/621)	0.034	(229/6641)	3.26	1.6×10^{-16}
DQB1*02:02	0.080	(78/974)	0.186	(2468/13273)	0.38	5.0×10^{-16}	0.096	(34/353)	0.193	(1282/6632)	0.44	6.5×10^{-6}	0.071	(44/621)	0.179	(1186/6641)	0.35	4.2×10^{-11}
DRB4*01:01	0.094	(92/974)	0.188	(2491/13273)	0.45	1.2×10^{-12}	0.113	(40/353)	0.194	(1289/6632)	0.52	0.0001	0.084	(52/621)	0.181	(1202/6641)	0.42	5.1×10^{-9}
B*44:03	0.038	(37/974)	0.105	(1397/13273)	0.34	2.8×10^{-10}	0.054	(19/353)	0.110	(728/6632)	0.46	0.001	0.029	(18/621)	0.101	(669/6641)	0.28	1.6×10^{-7}
C*16:01	0.030	(29/974)	0.082	(1090/13273)	0.35	2.9×10^{-8}	0.042	(15/353)	0.086	(573/6632)	0.46	0.004	0.023	(14/621)	0.078	(517/6641)	0.28	4.2×10^{-6}
							Diffuse SSc vs. Control											
HLA Allele	**dcSSc Prop.(Count)**		**CO Prop.(Count)**		**OR**	**p**	**dcSSc Prop.(Count)**		**CO prop.(Count)**		**OR**	***p***	**dcSSc prop.(Count)**		**CO prop. Count)**		**OR**	***p***
DPB1*13:01	0.108	(51/474)	0.033	(444/13273)	3.20	2.4×10^{-13}	0.109	(15/137)	0.035	(231/6632)	3.02	0.0001	0.107	(36/337)	0.032	(213/6641)	3.41	1.9×10^{-10}
DRB1*11:04	0.131	(62/474)	0.035	(459/13273)	2.93	3.3×10^{-12}	0.168	(23/137)	0.035	(230/6632)	3.47	1.9×10^{-6}	0.116	(39/337)	0.034	(229/6641)	2.71	2.6×10^{-7}
							Diffuse SSc vs. Limited SSc											
HLA Allele	**lcSSc Prop.(Count)**		**dcSSc prop.(Count)**		**OR**	***p***	**lcSSc prop.(Count)**		**dcSSc Prop.(Count)**		**OR**	**p**	**lcSSc prop.(Count)**		**dcSSc prop.(Count)**		**OR**	***p***
DQB1*02:02	0.08	(78/974)	0.146	(69/474)	2.04	6.2×10^{-5}	0.096	(34/353)	0.124	(17/137)	1.39	0.31	0.071	(44/621)	0.154	(52/337)	2.54	2.4×10^{-5}
DRB1*07:01	0.126	(123/974)	0.205	(97/474)	1.83	6.2×10^{-5}	0.150	(53/353)	0.175	(24/137)	1.24	0.44	0.113	(70/621)	0.217	(73/337)	2.23	1.4×10^{-5}
DQA1*02:01	0.127	(124/974)	0.205	(97/474)	1.81	8.0×10^{-5}	0.153	(54/353)	0.175	(24/137)	1.22	0.47	0.113	(70/621)	0.217	(73/337)	2.23	1.4×10^{-5}
							ACA + vs. –											
Allele	**ACA– prop.(Count)**		**ACA + prop.(Count)**		**OR**	***P***	**ACA– prop.(Count)**		**ACA + prop.(Count)**		**OR**	***P***	**ACA– prop.(Count)**		**ACA + prop.(Count)**		**OR**	***P***
DQA1*01:01	0.233	(210/903)	0.436	(221/507)	2.58	4.5×10^{-15}	0.241	(64/266)	0.401	(87/217)	2.21	0.0001	0.229	(146/637)	0.462	(134/290)	2.99	1.2×10^{-12}
DQB1*05:01	0.204	(184/903)	0.379	(192/507)	2.39	3.2×10^{-12}	0.214	(57/266)	0.35	(76/217)	2.02	0.0009	0.199	(127/637)	0.400	(116/290)	2.76	1.8×10^{-10}
DQA1*02:01	0.196	(177/903)	0.077	(39/507)	0.34	6.7×10^{-9}	0.207	(55/266)	0.106	(23/217)	0.45	0.004	0.192	(122/637)	0.055	(16/290)	0.24	3.1×10^{-7}
DRB1*07:01	0.195	(176/903)	0.077	(39/507)	0.34	9.3×10^{-9}	0.203	(54/266)	0.106	(23/217)	0.47	0.006	0.192	(122/637)	0.055	(16/290)	0.24	3.1×10^{-7}
DRB1*01:01	0.159	(144/903)	0.292	(148/507)	2.15	1.5×10^{-8}	0.162	(43/266)	0.281	(61/217)	2.04	0.002	0.159	(101/637)	0.300	(87/290)	2.33	8.7×10^{-7}
DQB1*02:02	0.134	(121/903)	0.043	(22/507)	0.29	2.8×10^{-7}	0.147	(39/266)	0.055	(12/217)	0.35	0.003	0.129	(82/637)	0.034	(10/290)	0.23	2.1×10^{-5}
							ATA + vs. –											
Allele	**ATA– prop.(Count)**		**ATA + prop.(Count)**		**OR**	***p***	**ATA– prop.(Count)**		**ATA + prop.(Count)**		**OR**	***p***	**ATA– prop.(Count)**		**ATA + prop.(Count)**		**OR**	***p***
DPB1*13:01	0.041	(48/1170)	0.250	(60/240)	7.85	9.3×10^{-22}	0.035	(14/404)	0.228	(18/79)	8.54	1.2×10^{-7}	0.044	(34/766)	0.261	(42/161)	8.39	6.8×10^{-16}
DRB1*11:04	0.083	(97/1170)	0.300	(72/240)	4.13	9.7×10^{-14}	0.069	(28/404)	0.266	(21/79)	4.93	9.1×10^{-6}	0.09	(69/766)	0.317	(51/161)	3.9	3.3×10^{-9}
DQA1*01:01	0.339	(397/1170)	0.133	(32/240)	0.30	2.4×10^{-9}	0.347	(140/404)	0.114	(9/79)	0.23	0.0001	0.336	(257/766)	0.143	(23/161)	0.33	4.2×10^{-6}
DPA1*02:01	0.271	(317/1170)	0.454	(109/240)	2.33	1.0×10^{-8}	0.260	(105/404)	0.519	(41/79)	3.26	5.9×10^{-6}	0.277	(212/766)	0.422	(68/161)	2.17	2.8×10^{-5}
DRB1*15:01	0.203	(237/1170)	0.350	(84/240)	2.40	4.3×10^{-8}	0.223	(90/404)	0.367	(29/79)	2.25	0.003	0.192	(147/766)	0.342	(55/161)	2.37	1.6×10^{-5}

(Continued on following page)

TABLE 3 | (*Continued*) HLA associations with SSc disease and major autoantibody subgroups.

HLA allele	Metanalysis				Cohort 1				Cohort 2			
					ATA + vs −							
	ATA− Prop.(Count)	ATA+ prop.(Count)	OR	*p*	ATA− Prop.(Count)	ATA+ prop.(Count)	OR	*p*	ATA− Prop.(Count)	ATA+ prop.(Count)	OR	*p*
DRB5*01:01	0.205 (240/1170)	0.354 (85/240)	2.38	5.0×10^{-8}	0.223 (90/404)	0.367 (29/79)	2.25	0.003	0.196 (150/766)	0.348 (56/161)	2.33	1.8×10^{-5}
DQB1*06:02	0.197 (231/1170)	0.329 (79/240)	2.33	2.2×10^{-7}	0.210 (85/404)	0.367 (29/79)	2.55	0.0008	0.191 (146/766)	0.311 (50/161)	2.11	0.0002
					ANoA + vs. −							
Allele	ANoA− prop.(Count)	ANoA + prop.(Count)	OR	*p*	ANoA− prop.(Count)	ANoA + prop.(Count)	OR	*p*	ANoA− prop.(Count)	ANoA + prop.(Count)	OR	*p*
DQA1*01:01	0.340 (359/1055)	0.201 (71/353)	0.48	9.2×10^{-7}	0.335 (119/355)	0.244 (31/127)	0.62	0.04	0.343 (240/700)	0.177 (40/226)	0.41	4.6×10^{-6}
DQA1*05:01	0.474 (500/1055)	0.626 (221/353)	1.84	1.6×10^{-6}	0.473 (168/355)	0.622 (79/127)	1.84	0.005	0.474 (332/700)	0.628 (142/226)	1.83	0.0001
DQB1*02:01	0.227 (240/1055)	0.337 (119/353)	1.88	4.7×10^{-6}	0.262 (93/355)	0.339 (43/127)	1.47	0.09	0.21 (147/700)	0.336 (76/226)	2.21	6.4×10^{-6}
C*07:01	0.304 (321/1055)	0.433 (153/353)	1.78	5.8×10^{-6}	0.332 (118/355)	0.457 (58/127)	1.69	0.01	0.29 (203/700)	0.42 (95/226)	1.86	0.0001

CO = control, OR = odds ratio, p = p-value, Prop. = proportion.

almost exclusively by ATA + patients, in which these alleles occurred at significantly higher frequency than in ATA − disease (*HLA-DPB1*13:01* OR = 7.85, $p_{FDR} = 9.3 \times 10^{-22}$; *HLA-DRB1*11:04* OR = 4.13, $p = 9.7 \times 10^{-14}$), increasing odds of ATA + SSc disease specifically substantially above that of SSc alone (**Table 3**). *HLA-DPB1*13:01* was seen in 25 and 4%, and *HLA-DRB1*11:04* in 30 and 8% of patients positive and negative for this autoantibody respectively. A significant increase in the frequency of alleles *HLA-DPA1*02:01* (OR = 2.33, $p = 1.0 \times 10^{-8}$), *DRB1*15:01* (OR = 2.40, $p = 4.3 \times 10^{-8}$) and *DRB5*01:01* (OR = 2.38, $p = 5.0 \times 10^{-8}$) was also seen in ATA + relative to ATA − SSc, though these alleles did not show a significant association with risk of SSc, or either disease subtype, overall. The strongest HLA class I associations were seen with lcSSc rather than dcSSc (*HLA-B*44:03* OR = 0.34, $p = 2.8 \times 10^{-10}$; *HLA-C*16:01* OR = 0.35, $p = 2.9 \times 10^{-8}$ lcSSc versus control; **Table 3**). Extended tables of HLA allele frequencies in each antibody positive subgroup of disease are provided in **Supplementary Tables S7–S9**.

KIR-HLA Class I Interactions in SSc

Given that certain HLA class I alleles serve as ligands for inhibitory and activating KIRs, an interaction evolved to buffer the innate killing activity of NK cells and other lymphocyte populations, HLA class I subgroup and allele associations with SSc were assessed in the context of genetic epistasis with their canonical KIRs. *KIR* imputation approaches allow profiling and statistical analysis of the highly polymorphic *KIR* locus in sizeable cohorts unamenable to lab-based copy-number typing. Imputed gene and haplotype frequencies across both study cohorts were in close agreement with data from publicly available and published European cohorts (**Supplementary Figures S3,S4**) and utilised for statistical comparison between SSc cases and controls. Before accounting for HLA background, none of the 14 KIR loci showed a significant association with SSc under a dominant or recessive inheritance model. SSc associated allele HLA-B*44:03, protective for lcSSc, is a member of the HLA-Bw4 subclass of KIR ligands recognised by inhibitory receptor KIR3DL1. The linked-allele HLA-C*16:01 is a HLA-C1 ligand carrying a residue 80 asparagine recognised by inhibitory receptors KIR2DL2 and KIR2DL3. An epistatic interaction was detected between KIR3DL1 and HLA-Bw4 in the SSc meta-analysis cohort, such that fewer KIR3DL1+ SSc patients carried an appropriate HLA-Bw4 ligand for this inhibitory receptor (67.7%) relative to KIR3DL1+ controls (72.2%, $p_{int} = 0.02$); accordingly, the protective association of HLA-Bw4 alleles with SSc was seen only in KIR3DL1+ individuals (OR = 0.77, $p = 2.0 \times 10^{-5}$; **Table 4**). No significant difference was observed in HLA-B*44 carriage in KIR3DL1+ compared with KIR3DL1-carriers, although the very small sample sizes preclude reliable interpretation. An interaction was also observed between HLA-C1 ligands and KIR2DL3/L2/S2 genes (all of which show perfect LD), such that HLA-C1 individuals carrying the activating receptor KIR2DS2 (and thus lacking paired inhibitory receptor KIR2DL3 in inverse LD) were at an increased risk of disease relative to KIR2DL3+ individuals ($p_{int} = 0.03$). The inhibitory receptor KIR2DL3 was seen at higher frequency in HLA-C*16 +

TABLE 4 | KIR interactions with HLA class I subtypes and alleles.

–		Metanalysis					Cohort 1					Cohort 2				
HLA	**KIR**	**SSc prop.(Count)**	**CO prop.(Count)**	**OR**	***p***	**p_{int}**	**SSc prop.(Count)**	**CO prop.(Count)**	**OR**	***p***	**p_{int}**	**SSc Prop.(Count)**	**CO prop.(Count)**	**OR**	***p***	**p_{int}**
Bw4	*3DL1+*	0.677 (947/1399)	0.722 (9121/12639)	0.77	$2x10^{-5}$	0.02	0.683 (332/486)	0.72 (4542/6305)	0.81	0.03	0.06	0.674 (615/913)	0.723 (4579/6334)	0.75	0.0002	0.15
	3DL1-	0.833 (55/66)	0.729 (462/634)	1.62	0.17		0.941 (16/17)	0.722 (236/327)	5.21	0.12		0.796 (39/49)	0.736 (226/307)	1.27	0.54	
C1	*2DL3+*	0.861 (1139/1323)	0.871 (10537/12097)	0.99	0.88	0.03	0.889 (399/449)	0.87 (5270/6058)	1.29	0.10	0.60	0.847 (740/874)	0.872 (5267/6039)	0.86	0.15	0.02
	2DL3-/L2++/S2++	0.93 (132/142)	0.867 (1020/1176)	2.14	0.03		0.926 (50/54)	0.88 (505/574)	1.96	0.22		0.932 (82/88)	0.855 (515/602)	2.63	0.03	
		Metanalysis					**Cohort 1**					**Cohort 2**				
KIR	**HLA Allele**	**SSc prop.(Count)**	**CO prop.(Count)**	**OR**	***p***	**p_{pnt}**	**SSc prop.(Count)**	**CO prop.(Count)**	**OR**	***p***	**p_{int}**	**SSc prop.(Count)**	**CO prop.(Count)**	**OR**	***p***	**p_{int}**
2DL3	*C*16+ (C1/C2)*	0.797 (59/74)	0.902 (1044/1157)	0.44	0.009	0.02	0.828 (24/29)	0.902 (543/602)	0.49	0.18	0.43	0.778 (35/45)	0.903 (501/555)	0.38	0.02	0.01
	*C*16- (C1/C2)*	0.909 (1264/1391)	0.912 (11053/12116)	1.01	0.93		0.897 (425/474)	0.915 (5515/6030)	0.85	0.30		0.915 (839/917)	0.910 (5538/6086)	1.12	0.37	

Statistical interactions between biologically interacting KIR and HLA subgroups or alleles reaching statistical significance (p < 0.05) in the meta-analysis cohort. + = positive, − = negative, ++ = homozygous, OR = odds ratio, CO = control, p = p-value, pint = KIRxHLA interaction term p-value.

controls (90.2%) than HLA-C*16 + SSc, cases (79.7%; OR = 0.44, $p = 0.009$), though did not differ in frequency in HLA-C*16- case-control analysis (p_{int} = 0.02; **Table 4**), suggesting that the protective association of this class I allele may be mediated through inhibitory KIR, interactions. Extended tables showing statistical interactions between biologically validated KIR-HLA, pairs are provided in **Supplementary Tables S10–S11**.

DISCUSSION

Both protective and disease-risk associations with classical HLA alleles are among the strongest seen in human immune-mediated disease, attesting to the central role of these molecules in dictating productive and pathologic immune responses. Additional to their function in guiding T cell specificity through presentation of self and foreign derived antigens, engagement of HLA ligands in a less antigen-dependent manner by inhibitory and activating lymphocyte receptors make them key moderators of lymphocyte activity. Systemic sclerosis remains an immunological perplexity; strong HLA associations, alongside emergence of serum autoantibodies that segregate with clinical phenotype, suggest an autoimmune axis. However, whether autoantibodies are drivers of cascading inflammation, fibrosis, organ destruction and resulting mortality in this disease remains unknown. Here we report on HLA associations detected in two independent and meta-analysed cohorts of SSc patients, and present findings from direct case-case comparisons of HLA frequencies in autoantibody positive and negative subgroups of disease. Furthermore, we identify statistical interactions between KIRs and HLA class I ligands in the largest analysis of genetic epistasis between these loci in SSc.

Meta-analysis of HLA frequencies in SSc patients and healthy controls revealed strong risk associations with class II alleles *HLA-DRB1*11:04* and *HLA-DPB1*13:01*, confirmed by step-wise and pair-wise conditional analysis to represent two independent disease-associated loci. *HLA-DRB1*11:04* associations were seen with both lcSSc and dcSSc, but the *HLA-DPB1*13:01* association reached GWS in the dcSSc cohort alone. The frequencies of both *HLA-DPB1*13:01* and *HLA-DRB1*11:04* alleles were significantly increased in ATA + relative to ATA −disease (OR = 7.85 and OR = 4.13 respectively), however neither differed in frequency from controls in ATA − or ACA + SSc cohorts, indicating the strong specificity of these genetic association for patients with ATA + serology. Both *HLA-DRB1*11:04* and *HLA-DPB1*13:01* alleles have been reported across a diversity of ethnic groups to be associated with SSc risk, and, specifically, ATA + disease (Zochling et al., 2014; Gourh et al., 2020; Acosta-Herrera et al., 2021). Furthermore, *HLA-DRB1*11* positivity, alongside ATAs and/or dcSSc, has been reported as the strongest risk factor for SSc associated pulmonary fibrosis in a United Kingdom cohort (Fanning et al., 1998). In a recent functional investigation of the link between topoisomerase-I (TOP1) derived peptides and *HLA-DR* alleles in 6 ATA + SSc patients, a restricted and shared set of TOP1 peptides were eluted off HLA-DR molecules from patient monocyte-derived dendritic cells pulsed with recombinant human TOP1. Sequence analysis of

ATA + SSc associated *HLA-DR* alleles (*HLA-DRB1*11:01*, **11:04*, **08:02*, **08:04,* and**15:02*), which were inherited by 3 of the 6 patients, revealed two strongly conserved motifs in the peptide-contact region of the DRβ-chain. Two further individuals carried *HLA-DRB1*15:01* and **16:01*, alleles with no previously reported ATA + SSc association, but which contained identical peptide-contact motifs to ATA−associated alleles, and shared TOP1 epitopes isolated from all such alleles elicited $CD4^+$ T cell activation in 73% of ATA + SSc patients relative to 27% of ATA−patients (Tiniakou et al., 2020). Relevant to this, this study is the first to show a GWS association of *HLA-DRB1*15:01* with ATA + relative to ATA − SSc (OR = 2.4, $p = 4.3 \times 10^{-8}$), following the recently published association of this allele with ATA + SSc by case-control analysis (Acosta-Herrera et al., 2021), suggesting commonalities in the preferential binding of this and other ATA + SSc associated HLA-DRB1 alleles to TOP1 derived peptides. These findings support the hypothesis that exposure to an environmentally derived TOP1 peptide mimic, presented by risk class II alleles, triggers development of ATA + antibodies, and suggest that autoantibodies may play a direct role in SSc-associated lung disease.

Alleles *HLA-DRB1*07:01*, *DQA1*02:01, DQB1*02:02,* and *DRB4*01:01* demonstrated a protective association with SSc. Pairwise conditioning on either allele *HLA-DRB1*07:01* or *DQA1*02:01* was sufficient to abolish disease association at the remaining three loci, confirming tight linkage of these two lead disease-protective alleles on a common haplotype. This is in contrast to previous conditional analyses which have reported *HLA-DQB1*02:02* as carrying the strongest independent protective association with disease (Gourh et al., 2020; Acosta-Herrera et al., 2021). Here, the *DRB1*07:01-DQA1*02:01* haplotype was associated with lcSSc alone, alongside *HLA-DQB1*02:02,* showing a reduced frequency in lcSSc when compared to dcSSc directly ($p < 8 \times 10^{-5}$). The haplotype showed a GWS association with ACA + but not ATA + disease, as previously reported (Gourh et al., 2020; Acosta-Herrera et al., 2021).

To our knowledge, no studies have yet reported GWS HLA associations between auto-antibody positive and negative subgroups of SSc patients. In addition to associations with ATA + SSc discussed above, significantly increased frequencies of the class II alleles *HLA-DQA1*01:01, DQB1*05:01,* and *DRB1*01:01* were seen with ACA + relative to ACA − disease, confirming published findings from smaller studies comparing patient subgroups across multiple ethnicities (Kuwana et al., 1995; Fanning et al., 1998; Simeon et al., 2009). Recent mapping of antigen targets for serum autoantibodies taken from patients with SSc, Sjögren's syndrome and primary biliary cholangitis showed broad spectrum reactivity of ACAs against multiple centromere antigens in a subset of patients, suggesting recognition of the tertiary structure of the centromere-kinetochore macrocomplex rather than single protein derived epitopes (Kajio et al., 2021). Such complexity in the epitope mapping of ACAs is likely to also complicate derivation of linear peptides that may link ACA−associated class II alleles to autoantibody production, with binding of centromere derived peptides to genetically implicated HLAs yet to be shown. Alternate strategies such as investigating peptides binding and/or eluted from disease associated compared with protective alleles offers a different and potentially more productive approach.

We are the first to report GWS protective HLA class I association with alleles *HLA-B*44:03* and *HLA-C*16:01* in SSc, detected in both the total SSc cohort meta-analysis and, specifically, with lcSSc. Consistent with our findings, a suggestive HLA-B*44:03 association with the same direction of effect has been reported previously in European Americans; however this was lost upon conditioning with associated alleles *HLA-DQB1*02:02*, *DPB1*13:01,* and *DRB1*11:04* (Gourh et al., 2020). Here, pairwise and iterative conditional analysis demonstrated strong LD between *HLA-B*44:03* and *HLA-C*16:01*, and weaker but detectable LD with the protective class II haplotype containing *HLA-DRB1*07:01*. Although associations with both HLA class I alleles were weakened upon correction for all three independently associated class II alleles, *HLA-DRB1*11:04, DRB1*07:01* and *DPB1*13:01*, the genetic signal remained significant (p <0.0008), suggesting that the observed HLA class I associations may be more than accessory to co-inherited class II alleles. Of relevance, an epistatic interaction was observed between *HLA-C*16* and the KIR locus, suggesting this allele may play an independent yet complementary role in disease protection. Of *HLA-C*16* + individuals, 79.7% of SSc patients co-inherited *KIR2DL3* relative to 90.2% of controls (OR = 0.44, p = 0.009), whereas *KIR2DL3* frequencies did not differ in *HLA-C*16-* individuals. *KIR2DL3* encodes an inhibitory receptor for HLA-C1 subgroup ligands. In a functional investigation of KIR2DL3 binding to a panel of 97 bead-bound HLA class I alleles expressing a broad repertoire of peptides, HLA-C*16:01 (a C1 ligand) ranked 8th in its binding avidity for KIR2DL3 and was one of the strongest inhibitors of KIR2DL3+ NK cell degranulation (Moradi et al., 2021). In our analyses, KIR2DL3 was conversely associated with increased disease risk when inherited with HLA-C2 allele *HLA-C*04*, which binds the KIR2DL3 receptor with only 10% of the avidity seen for its strongest HLA-C1 binding partner (Moradi et al., 2021). These findings emphasise the importance of functional context when interpreting *KIR-HLA* epistatic interactions in disease, and suggest that co-inheritance of *HLA-C*16* and *KIR2DL3* may protect from SSc through increasing the activation threshold of KIR2DL3 expressing lymphocytes.

The potential contribution of KIR-HLA ligand interactions to the risk and/or severity of immune-mediated disease has been explored across many conditions. In our analyses, co-inheritance of *HLA-Bw4* ligands in the presence of their compatible inhibitory receptor *KIR3DL1* was significantly associated with disease protection in both independent disease cohorts (meta-analysis $p = 2 \times 10^{-5}$), and inheritance of HLA-C1 alleles in the absence of their inhibitory receptor *KIR2DL3* was significantly associated with disease risk, albeit with more modest strength of association (p = 0.03). This suggests that signalling through inhibitory KIRs may reduce the risk of chronic, damaging, immune activation and protect against SSc. Among other diseases with immune involvement, co-inheritance of *KIR3DL1* and *HLA-Bw4*, or specific *HLA-Bw4* alleles, has been associated with protection from multiple sclerosis (Hollenbach et al., 2016) and ankylosing spondylitis (Hanson et al., 2020), and both *KIR3DL1-HLA-Bw4* and *KIR2DL3-HLA-C1* co-inheritance

with protection from autoimmune hepatitis (Littera et al., 2016). An absence of inherited strong inhibitory KIR-HLA interacting partners may be hypothesised to predispose to a lower threshold of lymphocyte activation and increased risk of autoinflammatory disease. A study of NK cell frequency and function in SSc patients' blood demonstrated a significant increase in NK cell numbers in dcSSc relative to controls, and increased expression of cell surface activation markers CD16 and CD69 on both lcSSc and dcSSc NK populations (Horikawa et al., 2005). However, upon *in vitro* stimulation, SSc patient NK cells also exhibited a significant reduction in interferon γ (IFNγ) production, a reduced ability to lyse target cells, and reduced granzyme B secretion. Conflicting studies surround the functional role of NK cells in SSc, emphasising that the contribution of this lymphocyte subset to disease is poorly understood. Our analyses suggest that variable inheritance of NK inhibitory and activating receptors imparts an additional layer of genetic heterogeneity to immune cell involvement in this condition.

Here we report an extensive analysis of class I and II HLA associations with SSc, and clinical and serological subtypes of disease, in two independent and meta-analysed cohorts of patients and controls. The substantial size of our study cohort has enabled us to identify HLA associations at GWS that differentiate autoantibody positive and negative SSc patients, emphasising the genetic heterogeneity underpinning this disease. Furthermore, we identify two new HLA class 1 associations, and show that co-inheritance of HLA class I ligands and KIRs, receptors that serve as key modulators of lymphocyte activation, may further contribute to an individual's underlying risk of developing SSc. Clear elucidation of genetic associations with disease risk and autoantibody positivity in SSc may aid in functional studies addressing the inflammatory triggers for disease, in the form of both endogenous host antigens and environmental stimuli that disrupt immune tolerance.

DATA AVAILABILITY STATEMENT

Extended summary statistics for HLA and KIR association analyses are provided in the **Supplementary Material**. Access requests for participant genotypes should be made through the corresponding author MB matt.brown@kcl.ac.uk.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the local ethics committees at participating centres that recruited SSc patients and healthy controls for analysis. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JS, G-SN, JR, JW, WS, MN, SA, SP, MM, and MB contributed to SSc patient and healthy control recruitment and clinical data collection. MM and MB facilitated genotyping. Data pre-processing, imputation, statistical analysis and manuscript writing was performed by AH, with guidance in study design and statistical analysis provided by TK and MB. All authors contributed to manuscript drafting.

FUNDING

This work was in part supported by the National Health and Medical Research Council (NHMRC) Grant GNT1162767. The Australian Scleroderma Cohort Study is supported by Scleroderma Victoria, Scleroderma Australia, Arthritis Australia, Australian Rheumatology Association, philanthropic donations, Janssen, and GSK. MN is supported by an NHMRC Investigator Grant (GNT1176538).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.913196/full#supplementary-material

REFERENCES

Acosta-Herrera, M., Kerick, M., Lopez-Isac, E., Assassi, S., Beretta, L., Simeon-Aznar, C. P., et al. (2021). Comprehensive Analysis of the Major Histocompatibility Complex in Systemic Sclerosis Identifies Differential HLA Associations by Clinical and Serological Subtypes. *Ann. Rheum. Dis.* 80, 1040–1047. doi:10.1136/annrheumdis-2021-219884

Arnett, F. C., Gourh, P., Shete, S., Ahn, C. W., Honey, R. E., Agarwal, S. K., et al. (2010). Major Histocompatibility Complex (MHC) Class II Alleles, Haplotypes and Epitopes Which Confer Susceptibility or Protection in Systemic Sclerosis: Analyses in 1300 Caucasian, African-American and Hispanic Cases and 1000 Controls. *Ann. Rheumatic Dis.* 69, 822–827. doi:10.1136/ard.2009.111906

Assassi, S., Sharif, R., Lasky, R. E., Mcnearney, T. A., Estrada-Y-Martin, R. M., Draeger, H., et al. (2010). Predictors of Interstitial Lung Disease in Early Systemic Sclerosis: a Prospective Longitudinal Study of the GENISOS Cohort. *Arthritis Res. Ther.* 12, R166. doi:10.1186/ar3125

Atamas, S. P., Yurovsky, V. V., Wise, R., Wigley, F. M., Goter Robinson, C. J., Henry, P., et al. (1999). Production of Type 2 Cytokines by CD8+ Lung Cells Is Associated with Greater Decline in Pulmonary Function in Patients with Systemic Sclerosis. *Arthritis & Rheumatism* 42, 1168–1178. doi:10.1002/1529-0131(199906)42:6<1168::aid-anr13>3.0.co;2-l

Chewning, J. H., Gudme, C. N., Hsu, K. C., Selvakumar, A., and Dupont, B. (2007). KIR2DS1-Positive NK Cells Mediate Alloresponse Against the C2 HLA-KIR Ligand Group *In Vitro*. *J. Immunol.* 179, 854–868. doi:10.4049/jimmunol.179.2.854

Cortes, A., and Brown, M. A. (2011). Promise and Pitfalls of the Immunochip. *Arthritis Res. Ther.* 13, 101. doi:10.1186/ar3204

Cortes, A., Hadler, J., Pointon, J. P., Robinson, P. C., Karaderi, T., Leo, P., et al. (2013). Identification of Multiple Risk Variants for Ankylosing Spondylitis through High-Density Genotyping of Immune-Related Loci. *Nat. Genet.* 45, 730–738. doi:10.1038/ng.2667

Del Rio, A. P. T., Sachetto, Z., Sampaio-Barros, P. D., Marques-Neto, J. F., Londe, A. C. S., and Bertolo, M. B. (2013). HLA Markers for Poor Prognosis in Systemic Sclerosis Brazilian Patients. *Dis. Markers* 35, 73–78. doi:10.1155/2013/301415

Fanning, G. C., Welsh, K. I., Bunn, C., Du Bois, R., and Black, C. M. (1998). HLA Associations in Three Mutually Exclusive Autoantibody Subgroups in UK Systemic Sclerosis Patients. *Rheumatology* 37, 201–207. doi:10.1093/rheumatology/37.2.201

Fusco, C., Guerini, F. R., Nocera, G., Ventrella, G., Caputo, D., Valentino, M. A., et al. (2010). KIRs and Their HLA Ligands in Remitting-Relapsing Multiple Sclerosis. *J. Neuroimmunol.* 229, 232–237. doi:10.1016/j.jneuroim.2010.08.004

Gladman, D. D., Kung, T. N., Siannis, F., Pellett, F., Farewell, V. T., and Lee, P. (2005). HLA Markers for Susceptibility and Expression in Scleroderma. *J. Rheumatol.* 32, 1481–1487.

Gonzalez-Galarza, F. F., Mccabe, A., Santos, E. J. M. D., Jones, J., Takeshita, L., Ortega-Rivera, N. D., et al. (2020). Allele Frequency Net Database (AFND) 2020 Update: Gold-Standard Data Classification, Open Access Genotype Data and New Query Tools. *Nucleic Acids Res.* 48, D783–D788. doi:10.1093/nar/gkz1029

Gourh, P., Safran, S. A., Alexander, T., Boyden, S. E., Morgan, N. D., Shah, A. A., et al. (2020). HLA and Autoantibodies Define Scleroderma Subtypes and Risk in African and European Americans and Suggest a Role for Molecular Mimicry. *Proc. Natl. Acad. Sci. U.S.A.* 117, 552–562. doi:10.1073/pnas.1906593116

Gumperz, J. E., Valiante, N. M., Parham, P., Lanier, L. L., and Tyan, D. (1996). Heterogeneous Phenotypes of Expression of the NKB1 Natural Killer Cell Class I Receptor Among Individuals of Different Human Histocompatibility Leukocyte Antigens Types Appear Genetically Regulated, but Not Linked to Major Histocompatibililty Complex Haplotype. *J. Exp. Med.* 183, 1817–1827. doi:10.1084/jem.183.4.1817

Hanson, A. L., Leslie, S., Harris, J., Lê Cao, K.-A., Kenna, T. J., and Brown, M. A. (2020). Epistatic Interactions Between Killer Immunoglobulin-Like Receptors and Human Leukocyte Antigen Ligands Are Associated with Ankylosing Spondylitis. *PLoS Genet.* 16, e1008906. doi:10.1371/journal.pgen.1008906

Ho, K. T., and Reveille, J. D. (2003). The Clinical Relevance of Autoantibodies in Scleroderma. *Arthritis Res. Ther.* 5, 80–93. doi:10.1186/ar628

Hollenbach, J. A., Pando, M. J., Caillier, S. J., Gourraud, P.-A., and Oksenberg, J. R. (2016). The Killer Immunoglobulin-Like Receptor KIR3DL1 in Combination with HLA-Bw4 Is Protective Against Multiple Sclerosis in African Americans. *Genes Immun.* 17, 199–202. doi:10.1038/gene.2016.5

Horikawa, M., Hasegawa, M., Komura, K., Hayakawa, I., Yanaba, K., Matsushita, T., et al. (2005). Abnormal Natural Killer Cell Function in Systemic Sclerosis: Altered Cytokine Production and Defective Killing Activity. *J. Investigative Dermatology* 125, 731–737. doi:10.1111/j.0022-202x.2005.23767.x

Jacobsen, S., Ullman, S., Shen, G. Q., Wiik, A., and Halberg, P. (2001). Influence of Clinical Features, Serum Antinuclear Antibodies, and Lung Function on Survival of Patients with Systemic Sclerosis. *J. Rheumatol.* 28, 2454–2459.

Jandali, B., Salazar, G. A., Hudson, M., Fritzler, M. J., Lyons, M. A., Estrada, Y. M. R. M., et al. (2022). The Effect of Anti-Scl-70 Antibody Determination Method on its Predictive Significance for Interstitial Lung Disease Progression in Systemic Sclerosis. *ACR Open Rheumatol.* 4, 345–351. doi:10.1002/acr2.11398

Jiang, W., Johnson, C., Jayaraman, J., Simecek, N., Noble, J., Moffatt, M. F., et al. (2012). Copy Number Variation Leads to Considerable Diversity for B but Not A Haplotypes of the Human KIR Genes Encoding NK Cell Receptors. *Genome Res.* 22, 1845–1854. doi:10.1101/gr.137976.112

Kajio, N., Takeshita, M., Suzuki, K., Kaneda, Y., Yamane, H., Ikeura, K., et al. (2021). Anti-Centromere Antibodies Target Centromere-Kinetochore Macrocomplex: A Comprehensive Autoantigen Profiling. *Ann. Rheum. Dis.* 80, 651–659. doi:10.1136/annrheumdis-2020-218881

Kettaneh, A., Al Moufti, O., Tiev, K. P., Chayet, C., Tolédano, C., Fabre, B., et al. (2007). Occupational Exposure to Solvents and Gender-Related Risk of Systemic Sclerosis: A Metaanalysis of Case-Control Studies. *J. Rheumatol.* 34, 97–103.

Khakoo, S. I., Thio, C. L., Martin, M. P., Brooks, C. R., Gao, X., Astemborski, J., et al. (2004). HLA and NK Cell Inhibitory Receptor Genes in Resolving Hepatitis C Virus Infection. *Science* 305, 872–874. doi:10.1126/science.1097670

Kuwana, M., Okano, Y., Kaburaki, J., and Inoko, H. (1995). HLA Class II Genes Associated with Anticentromere Antibody in Japanese Patients with Systemic Sclerosis (Scleroderma). *Ann. Rheumatic Dis.* 54, 983–987. doi:10.1136/ard.54.12.983

Leroy, E. C., and Medsger, T. A. (2001). Criteria for the Classification of Early Systemic Sclerosis. *J. Rheumatol.* 28, 1573–1576.

Leroy, E. C. (1974). Increased Collagen Synthesis by Scleroderma Skin Fibroblasts *In Vitro* A Possible Defect in the Regulation or Activation of the Scleroderma Fibroblast. *J. Clin. Invest.* 54, 880–889. doi:10.1172/jci107827

Li, G., Larregina, A. T., Domsic, R. T., Stolz, D. B., Medsger, T. A., Jr., Lafyatis, R., et al. (2017). Skin-Resident Effector Memory CD8+CD28- T Cells Exhibit a Profibrotic Phenotype in Patients with Systemic Sclerosis. *J. Investigative Dermatology* 137, 1042–1050. doi:10.1016/j.jid.2016.11.037

Littera, R., Chessa, L., Onali, S., Figorilli, F., Lai, S., Secci, L., et al. (2016). Exploring the Role of Killer Cell Immunoglobulin-Like Receptors and Their HLA Class I Ligands in Autoimmune Hepatitis. *PLoS One* 11, e0146086. doi:10.1371/journal.pone.0146086

Lunardi, C., Dolcino, M., Peterlana, D., Bason, C., Navone, R., Tamassia, N., et al. (2006). Antibodies Against Human Cytomegalovirus in the Pathogenesis of Systemic Sclerosis: A Gene Array Approach. *PLoS Med.* 3, e2. doi:10.1371/journal.pmed.0030002

Mahmoudi, M., Fallahian, F., Sobhani, S., Ghoroghi, S., Jamshidi, A., Poursani, S., et al. (2017). Analysis of Killer Cell Immunoglobulin-Like Receptors (KIRs) and Their HLA Ligand Genes Polymorphisms in Iranian Patients with Systemic Sclerosis. *Clin. Rheumatol.* 36, 853–862. doi:10.1007/s10067-016-3526-0

Maul, G. G., Jimenez, S. A., Riggs, E., and Ziemnicka-Kotula, D. (1989). Determination of an Epitope of the Diffuse Systemic Sclerosis Marker Antigen DNA Topoisomerase I: Sequence Similarity with Retroviral P30gag Protein Suggests a Possible Cause for Autoimmunity in Systemic Sclerosis. *Proc. Natl. Acad. Sci. U.S.A.* 86, 8492–8496. doi:10.1073/pnas.86.21.8492

Mayes, M. D., Bossini-Castillo, L., Gorlova, O., Martin, J. E., Zhou, X., Chen, W. V., et al. (2014). Immunochip Analysis Identifies Multiple Susceptibility Loci for Systemic Sclerosis. *Am. J. Hum. Genet.* 94, 47–61. doi:10.1016/j.ajhg.2013.12.002

Mierau, R., Moinzadeh, P., Riemekasten, G., Melchers, I., Meurer, M., Reichenberger, F., et al. (2011). Frequency of Disease-Associated and Other Nuclear Autoantibodies in Patients of the German Network for Systemic Scleroderma: Correlation with Characteristic Clinical Features. *Arthritis Res. Ther.* 13, R172. doi:10.1186/ar3495

Moradi, S., Stankovic, S., O'Connor, G. M., Pymm, P., Maclachlan, B. J., Faoro, C., et al. (2021). Structural Plasticity of KIR2DL2 and KIR2DL3 Enables Altered Docking Geometries Atop HLA-C. *Nat. Commun.* 12, 2173. doi:10.1038/s41467-021-22359-x

Motyer, A., Vukcevic, D., Dilthey, A., Donnelly, P., Mcvean, G., and Leslie, S. (2016). *Practical Use of Methods for Imputation of HLA Alleles from SNP Genotype Data*. NY.USA: bioRxiv.

O'connor, G. M., Vivian, J. P., Gostick, E., Pymm, P., Lafont, B. A. P., Price, D. A., et al. (2015). Peptide-Dependent Recognition of HLA-B*57:01 by KIR3DS1. *J. Virol.* 89, 5213–5221. doi:10.1128/jvi.03586-14

Pellett, F., Siannis, F., Vukin, I., Lee, P., Urowitz, M. B., and Gladman, D. D. (2007). KIRs and Autoimmune Disease: Studies in Systemic Lupus Erythematosus and Scleroderma. *Tissue antigens* 69, 106–108. doi:10.1111/j.1399-0039.2006.762_6.x

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reveille, J. D., and Solomon, D. H. (2003). Evidence-based Guidelines for the Use of Immunologic Tests: Anticentromere, Scl-70, and Nucleolar Antibodies. *Arthritis & Rheumatism* 49, 399–412. doi:10.1002/art.11113

Reveille, J. D., Fischbach, M., Mcnearney, T., Friedman, A. W., Aguilar, M. B., Lisse, J., et al. (2001). Systemic Sclerosis in 3 US Ethnic Groups: A Comparison of Clinical, Sociodemographic, Serologic, and Immunogenetic Determinants. *Seminars Arthritis Rheumatism* 30, 332–346. doi:10.1053/sarh.2001.20268

Salim, P. H., Jobim, M., Bredemeier, M., Chies, J. A. B., Schlottfeldt, J., Brenol, J. C. T., et al. (2010). Killer Cell Immunoglobulin-Like Receptor (KIR) Genes in Systemic Sclerosis. *Clin. Exp. Immunol.* 160, 325–330. doi:10.1111/j.1365-2249.2010.04095.x

Servaas, N. H., Zaaraoui-Boutahar, F., Wichers, C. G. K., Ottria, A., Chouri, E., Affandi, A. J., et al. (2021). Longitudinal Analysis of T-Cell Receptor Repertoires Reveals Persistence of Antigen-Driven CD4+ and CD8+ T-Cell Clusters in Systemic Sclerosis. *J. Autoimmun.* 117, 102574. doi:10.1016/j.jaut.2020.102574

Simeón, C. P., Fonollosa, V., Tolosa, C., Palou, E., Selva, A., Solans, R., et al. (2009). Association of HLA Class II Genes with Systemic Sclerosis in Spanish Patients. *J. Rheumatol.* 36, 2733–2736. doi:10.3899/jrheum.090377

Sluis-Cremer, G. K., Hessel, P. A., Nizdo, E. H., Churchill, A. R., and Zeiss, E. A. (1985). Silica, Silicosis, and Progressive Systemic Sclerosis. *Occup. Environ. Med.* 42, 838–843. doi:10.1136/oem.42.12.838

Steen, V. D., Powell, D. L., and Medsger, T. A., Jr. (1988). Clinical Correlations and Prognosis Based on Serum Autoantibodies in Patients with Systemic Sclerosis. *Arthritis & Rheumatism* 31, 196–203. doi:10.1002/art.1780310207

Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee (1980). Preliminary Criteria for the Classification of Systemic Sclerosis (Scleroderma). Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. *Arthritis Rheum.* 23, 581–590. doi:10.1002/art.1780230510

Swanson Beck, J., Anderson, J. R., Gray, K. G., and Rowell, N. R. (1963). Antinuclear and Precipitating Autoantibodies in Progressive Systemic Sclerosis. *Lancet* 282, 1188–1190. doi:10.1016/s0140-6736(63)92920-9

Tiniakou, E., Fava, A., Mcmahan, Z. H., Guhr, T., O'meally, R. N., Shah, A. A., et al. (2020). Definition of Naturally Processed Peptides Reveals Convergent Presentation of Autoantigenic Topoisomerase I Epitopes in Scleroderma. *Arthritis Rheumatol.* 72, 1375–1384. doi:10.1002/art.41248

Vukcevic, D., Traherne, J. A., Næss, S., Ellinghaus, E., Kamatani, Y., Dilthey, A., et al. (2015). Imputation of KIR Types from SNP Variation Data. *Am. J. Hum. Genet.* 97, 593–607. doi:10.1016/j.ajhg.2015.09.005

Winter, C. C., Gumperz, J. E., Parham, P., Long, E. O., and Wagtmann, N. (1998). Direct Binding and Functional Transfer of NK Cell Inhibitory Receptors Reveal Novel Patterns of HLA-C Allotype Recognition. *J. Immunol.* 161, 571–577.

Yurovsky, V. V., Wigley, F. M., Wise, R. A., and White, B. (1996). Skewing of the CDS + T-Cell Repertoire in the Lungs of Patients with Systemic Sclerosis. *Hum. Immunol.* 48, 84–97. doi:10.1016/0198-8859(96)00091-2

Zochling, J., Newell, F., Charlesworth, J. C., Leo, P., Stankovich, J., Cortes, A., et al. (2014). An Immunochip-Based Interrogation of Scleroderma Susceptibility Variants Identifies a Novel Association at DNASE1L3. *Arthritis Res. Ther.* 16, 438. doi:10.1186/s13075-014-0438-8

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

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

frontiers | Frontiers in Pediatrics

TYPE Case Report
PUBLISHED 09 March 2023
DOI 10.3389/fped.2023.1118097

OPEN ACCESS

EDITED BY
Jinqiao Sun,
Fudan University, China

REVIEWED BY
Tiphanie Phillips Vogel,
Baylor College of Medicine, United States
Daniel Calame,
Baylor College of Medicine, United States

*CORRESPONDENCE
Xiaolei Tang
✉ tangxl9@gmail.com

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

RECEIVED 07 December 2022
ACCEPTED 20 February 2023
PUBLISHED 09 March 2023

CITATION
Li X, Tang Y, Zhang L, Wang Y, Zhang W, Wang Y, Shen Y and Tang X (2023) Case report: COPA syndrome with interstitial lung disease, skin involvement, and neuromyelitis spectrum disorder.
Front. Pediatr. 11:1118097.
doi: 10.3389/fped.2023.1118097

Case report: COPA syndrome with interstitial lung disease, skin involvement, and neuromyelitis spectrum disorder

Xiao Li[1], Yu Tang[1], Lei Zhang[1], Yuan Wang[2], Weihua Zhang[3], Ying Wang[2], Yuelin Shen[4] and Xiaolei Tang[4]*

[1]Department of Respiratory Medicine, Children's Hospital Affiliated to Zhengzhou University/Henan Children's Hospital/Zhengzhou Children's Hospital, Zhengzhou, China, [2]Department of Neurology, Children's Hospital Affiliated to Zhengzhou University/Henan Children's Hospital/Zhengzhou Children's Hospital, Zhengzhou, China, [3]Department of Neurology, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, China, [4]Department of Respiratory Medicine, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, China

This report describes a case of a 22 months Chinese boy with COPA syndrome bearing the c.715G > C (p.A239P) genotype. In addition to interstitial lung diseae, he also suffered from recurrent chilblain-like rashes, which has not been previously reported, and neuromyelitis optica spectrum disorder (NMOSD), which is a very rare phenotype. Clinical manifestations expanded the phenotype of COPA syndrome. Notably, there is no definitive treatment for COPA syndrome. In this report, the patient has achieved short-term clinical improvement with sirolimus.

KEYWORDS

children, copa syndrome, neuromyelitis optica spectrum disorder, rashes, sirolimus

Introduction

In 2019, the International Union of Immunological Societies Expert Committee (IUIS) classified COPA syndrome as a non-inflammasome-associated disease (1). COPA syndrome shares very similar symptoms and pathogenesis with another Type I interferonopathy, namely, STING-associated vasculopathy with onset in infancy (SAVI). STING and subsequent IFN pathway are activated by the mutation of *STING1* gene. Up till now, a total of 78 individuals carrying 16 variants of *COPA* gene have been reported in the literature. COPA syndrome, also known as an immune deficiency disease with strong clinical heterogeneity and certain commonality (2), usually affects the lungs, kidneys, and joints. Skin vasculopathy is considered as a core feature of SAVI, which was observed in 77% of the reported patients, but has yet to be described in any patient with COPA syndrome. In this report, we present a pediatric case with recurrent chilblain-like rashes, which is a kind of skin vasculopathy. In addition, the patient also suffered from NMOSD, which has been reported only once before.

Abbreviations
NMOSD: neuromyelitis optica spectrum disorder; AQP4, anti-aquaporin 4 antibody; IFN, interferon; MRI, magnetic resonance imaging; pANCA, perinuclear antineutrophil cytoplasmic antibody; ANA, antinuclear antibody; MOG, anti-myelin oligodendrocyte glycoprotein antibody; GFAP, anti-glial fibrillary acidic protein; ILD, interstitial lung disease; SLE, systemic lupus erythematosus; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate.

10.3389/fped.2023.1118097

Case presentation

A 22-month-old boy presented with irritability for 8 days and limb weakness for 6 days, accompanied by lethargy and vomiting. The boy had a history of transient paralysis of the right upper limb after fever at the age of 19-month-old, and his symptoms alleviated three days later. Additionally, he also manifested repeated facial and ear rashes, which were diagnosed as chilblain. Physical examination showed decreased muscle strength (3+/5 and 3/5 for upper and lower limbs, respectively), mild hypertonia in the left lower limb, nuchal rigidity, normal deep reflex, negative Babinski's sign, Kernig's sign and Brudzinski's sign. Laboratory tests (**Table 1**) revealed positive perinuclear antineutrophil cytoplasmic antibody (pANCA) and antinuclear antibody (ANA) (1:320), negative anti-double-stranded DNA (ds-DNA) antibody and anti-Sm antibody, positive serum anti-aquaporin 4 antibody (AQP4) (1:1,000, cell based assay), negative serum anti-myelin oligodendrocyte glycoprotein antibody (MOG), anti-glial fibrillary acidic protein (GFAP) antibody and serum antibodies against autoimmune peripheral neuropathy (KingMed Diagnostics). No abnormalities was found in ophthalmic examination revealed. Routine and biochemical test of cerebrospinal fluid (CSF) were normal. CSF culture and metagenomic next-generation sequencing (mNGS) showed no etiology. Brain MRI showed lesions involving medulla and white matter around the ventricle and small encephalomalacia foci with gliosis in the left basal ganglia (**Figure 1A**). MRI of the optic nerve was normal (**Figure 1B**). MRI of the Spinal cord showed longitudinally extensive transverse myelitis lesions from the medulla to C6 level (**Figures 1C,D**). According to the NMOSD diagnostic criteria proposed by the NMO diagnostic Group in 2015 (2), the child was diagnosed as NMOSD (based on the existence of two core clinical symptoms, namely, acute myelitis and area postrema syndrome, and positive AQP4 antibody). The patient was reated with high dose of intravenous methylprednisolone (20 mg per kilogram for 4 day) followed by oral prednisone with gradual diminution of dose, along with intravenous immunogloblin (IVIG) (2 g/kg). Then his muscle strength gradually improved without recurrence. Disease-Modifying Treatment (DMT) is recommended as the standard treatment in remission of NMODS, which can reduce the clinical onset of patients with recurrent remission. Regrettably, his parents refused to use DMT in consideration of its adverse effects.

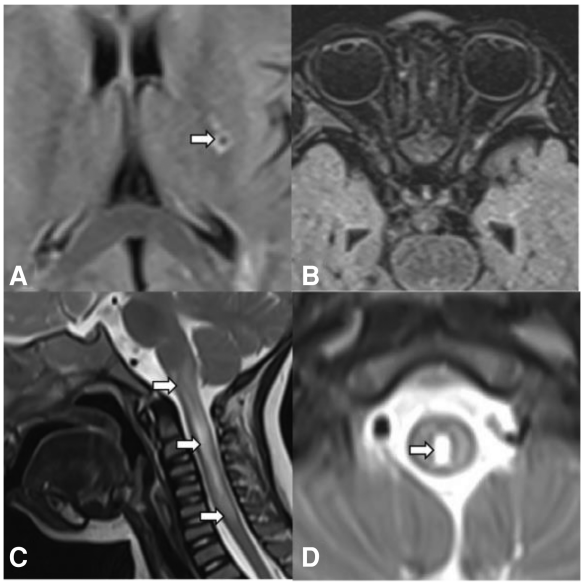


FIGURE 1
Brain MRI, spinal cord MRI, and optic MRI of the patient. Axial T2-weighted fluid-attenuated inversion recovery (FLAIR) MRI shows small encephalomalacia foci with gliosis in the left basal ganglia (A, arrows) and normal optic nerves (B). Sagittal (C) and axial (D) T2-weighted MRI of the spinal cord demonstrates longitudinally extensive transverse myelitis (LETM) lesion involving medulla to C6 leve (arrows).

TABLE 1 Clinical data of the patient with COPA syndrome.

Clinical data	
Sex	male
Age (m)	22
History	transient paralysis of the right upper limb
Symptom/sign	cough; tachypnea; clubbing; Chilblain - like rash; NMOSD
Lung	
CT chest	ground glass opacities (GGO); consolidations
TBFV	Mixed
Arthritis	–
Kidney	–
Allergic disease	Urticaria
Others	–
ANA	+ (1:320)
dsDNA	–
RF	+
HLA-B27	–
Anti-CCP	–
MPO	–
PR3	–
CRP (mg/L)	0.49 (0–10)
ESR (mm/h)	20 (1–15)
IL-6 (pg/ml)	3.4 (<7)
Total T lymphocytes	994.01 (770–2,860)
AQP4	+(1:1,000)

TBFV, tidal breathing flow volume curve; ANA, antinuclear antibody; CCP, cyclic citrullinated peptide; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; PFT, pulmonary function test; AQP4, anti-aquaporin 4 antibody.

The patient developed dry cough, shortness of breath accompanied with rashes in the following months. Physical examination showed tachypnea, retractions and clubbing fingers, as well as Chilblain-like rashes on both of cheeks (**Figure 2**). His respiratory symptoms revealed a lousy response to anti-infection treatment and deteriorated. Chest CT (**Figure 3**) showed ground glass opacities (GGO) and consolidations in both lungs, suggesting interstitial lung disease (ILD). The whole exome sequencing (WES), conducted by Beijing MyGenostics Laboratory, revealed a heterozygous variant of *COPA* gene, that is, c.715G > C (p.A239P, NM_004371), which was confirmed by sanger sequencing (**Figure 4**). The patient was administrated with prednisone 2 mg/kg per day for one month at the age of 25 months. His tachypnea and cough were slightly alleviated but not

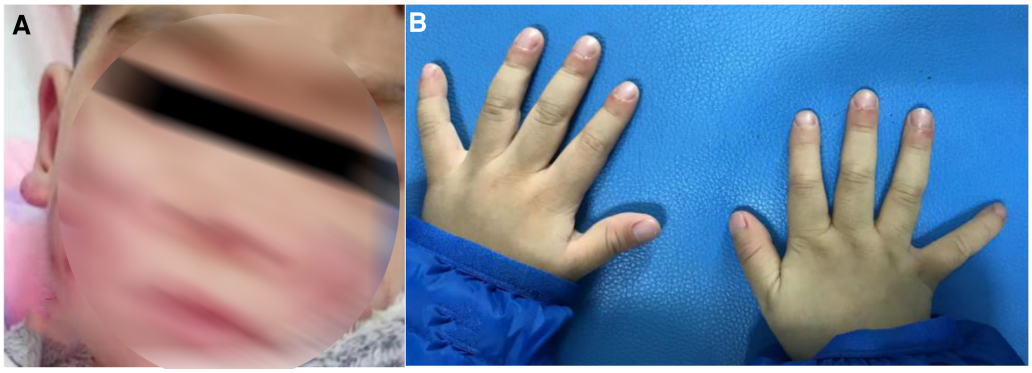


FIGURE 2
(A) Chilblain-like rash was observed on the face and ears of the child. (B) Clubbing fingers.

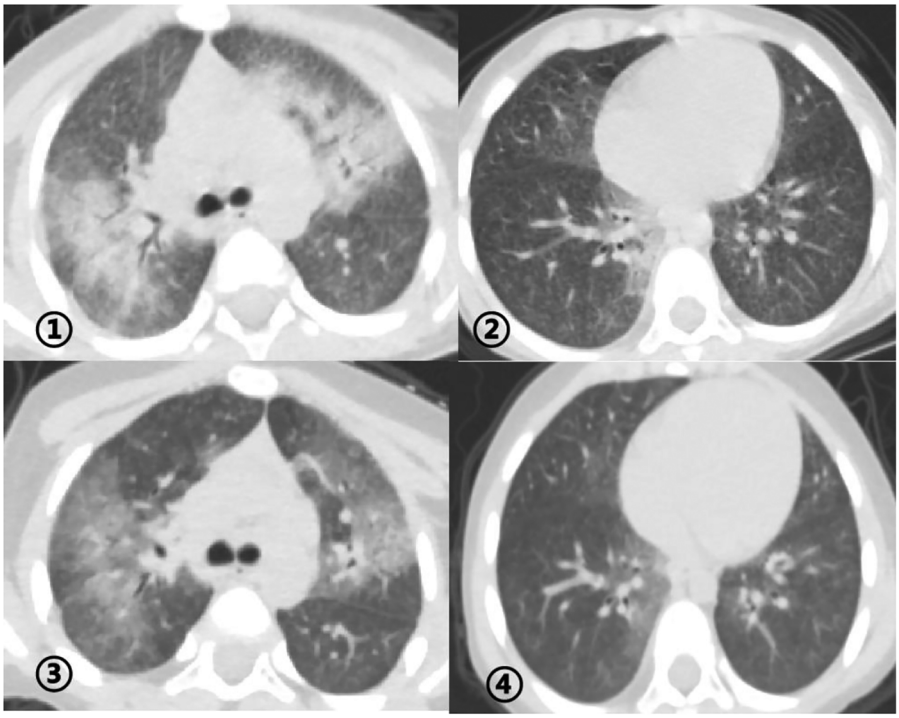


FIGURE 3
①②: Chest CT before the treatment at the age of 25-months old revealing ground glass opacities (GGO) and consolidations in both lungs. ③④: Chest CT at the age of 29-months old revealing significant absorption of GGO and consolidations in the lung after treatment for 4 months.

disappear during the treatment. Then, he received a combined treatment of sirolimus with a dose of 0.8 mg/m^2 at the age of 26 months. Meanwhile, the dosage of prednisone gradually decreased to 0.5 mg/kg and remained unchanged. Afterwards, his cough and tachypnea gradually disappeared. Repeated chest CT at the age of 29-months old showed significant absorption of GGO and consolidations in both lungs after the 4 months of prednisone treatment and 3 months of combined sirolimus treatment. However, the patient was non-adherent to follow up.

Discussion

COPA gene is located on chromosome1q23.2 and including 33 exons (54 kb). To date, most of the pathogenic COPA variants identified in patients with autoinflammatory diseases compatible with COPA syndrome are located in the 14 amino acid region and the WD40 domain with important functions, except for 4 mutations, that is c.433C > T (p.P145S) (3), c.596A > G (p.H199R) (4), c.841C > T (p.R281W) (5) and c.855G > C (p.Q285H) (6). The genetic testing of this patient revealed a variant of c.715G > C

10.3389/fped.2023.1118097

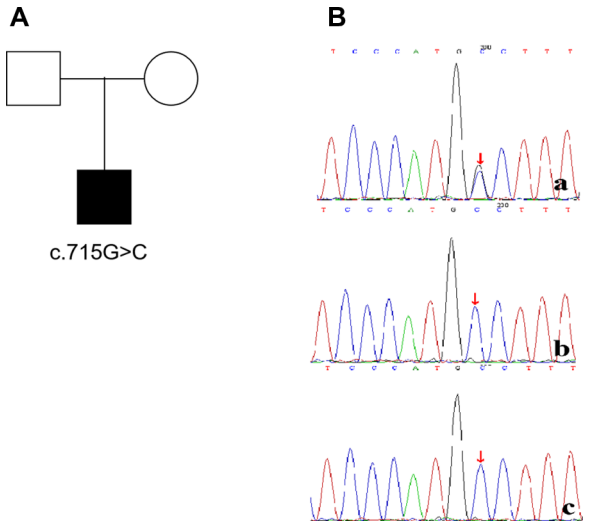


FIGURE 4
The child(a) has a *de novo* mutation (c.715G>C,p.A239P), his father(b) and mother(c) have no mutation at the same location.

(p.A239P), located within the WD40 domain, which has been previously reported by Pamela Psarianos (7).

Clinically, COPA syndrome mainly manifests diffuse alveolar hemorrhage or interstitial lung disease, arthritis, and renal injury. Additionally, its onset is mostly in childhood, without racial predilection. Watkin et al. (8) reported that all of the patients had lung disease diagnosed as pulmonary hemorrhage, interstitial lung disease, or both. Arthritis was found in 95% of children, with the knee joint and interphalangeal joints being the most common, and rheumatoid factor positive in 43% of children. In the available literature, only 4 patients have neurological symptoms (9). Bader-meunier B et al. (6) reported a case of COPA syndrome with arthritics only. The mutation occurred on the outer surface of the adjacent blade within the seven-bladed b-propellor structure, distinctting from the previously reported hot spot mutations. To our great knowledge, there have been no previous reports of rashes in patients with COPA syndrome.

In this case, the child presented with limb weakness, and MRI of the spinal cord indicated patchy abnormal signal shadows from C1 to C6, which was consistent with acute myelitis. In terms of his clinical sympotom of vomiting, encephalitis was excluded due to no fever and convulsive seizure, normal CSF routine and biochemistry, no etiology of CSF culture and mNGS, negative antibodies against autoimmune peripheral neuropathy. Brain MRI suggested abnormal signal shadow in the medulla oblongata, area postrema syndrome was therefore taken into consideration. At the same time, according to the NMOSD diagnostic criteria formulated by the International NMO Diagnostic Group (IPND) in 2015, the child could be diagnosed as NMOSD due to positive AQP4. Additionally, this child was transiently paralyzed in his right upper limb at the age of 19 months. Combined with head MRI, the left basal ganglia encephalomalacia lesion was shown (there was no change in subsequent brain MRI), which might be episode of focal cerebral ischemia. As far as we known, NMOSD is a very rare phenotype of COPA syndrome. In the past, only one patient suffered from hearing loss due to bacterial meningitis and bilateral recurrent neuromyelitis optica. However, the presence of AQP4-Ab was not reported in this patient (10). NMOSD is an uncommon antibody-mediated central nervous system disease. Antibodies against aquaporin-4 (AQP4-Ab), a water channel expressed on astrocytes has been found in approximately 75% of patients. A recent whole-genome sequence study identified genetic variants in the major histocompatibility region that contribute to the etiology of NMOSD (11). About one quarter of AQP4-Ab positive NMOSD patients suffer from another autoimmune disease, such as systemic lupus erythematosus (SLE) and myasthenia gravis (12, 13). SLE and NMOSD share a common origin of interferonopathy. Intriguingly, Jac Williams (14) reported that either an increase in endogenous IFNα (such as in patients with SLE) or an increase in exogenous IFNα (treatment with recombinant interferon alpha) could promote the development of NMOSD. COPA syndrome is an autoinflammatory disease with autoimmune features, suggesting that the pathogenesis of COPA syndrome and NMOSD comorbidity may be similar to the SLE and NMOSD comorbidities, which may be associated with humoral immunity.

Patients with COPA syndrome shares similar clinical features with SAVI, such as ILD, joint and kidney involvement (15). Skin vasculopathy, ranging from a mild rash or livedo to severe ulcerative lesions and extensive tissue loss, is a core feature of SAVI, which is observed in 77% of the reported patients, but has yet to be described in any previous COPA syndrome patients. This is the first case of chilblain-like rashes in COPA syndrome.

At present, COPA syndrome treatment mainly refers to other autoimmune diseases. Common treatment schemes include glucocorticoids in combination with immunosuppressors such as JAK1/2 inhibitor (ruxolitinib, baricitinib). JAK inhibitors are widely used to inhibit cytokine signaling, including the downstream of interferons and other cytokines. It is reported that the application of JAK inhibitors can improve the patient's well-being and quality of life of the patients (16, 17). Glucocorticoid combined with hydroxychloroquine and mycophenolate mofetil (MMF) can also alleviate cough in COPA patients (7). In addition, Guan Y et al. (2) reported that two cases of COPA syndrome were treated with sirolimus and achieved favourable therapeutic effect. In the present report, the symptoms of the child have been remarkably alleviated after the treatment of glucocorticoid in combination with sirolimus, which supports the possibility that sirolimus may serve as an effective treatment for COPA syndrome. Blocking the mTOR pathway may be the mechanism of sirolimus in treating COPA syndrome, which is one of the downstream signaling pathways activated by STING. The long-term prognosis of COPA syndrome remains unclear due to the limited number of COPA syndrome cases and insufficient follow-up periods.

There are still several limitations in our study. Skin biopsy was not performed to verify that the chilblain-like rash in the child was vasculitis. Moreover, we only conducted short-term follow-up.

Conclusion

In summary, we first reported a new phenotype of skin involvement in one patient with COPA syndrome, which expanded the phenotypic spectrum of this disease.

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 Children's Hospital Affiliated to Zhengzhou University/Henan Children's Hospital/Zhengzhou Children's Hospital. 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 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

XL and YW: supervised the patient care, conceptualized and designed the study, drafted the initial manuscript, and reviewed and revised the manuscript. YT and LZ: performed the lung scintigraphy, designed the study, and reviewed and revised the manuscript. YW and WZ: performed the nervous system scintigraphy, designed the study, and reviewed and revised the manuscript. YS: performed genetic analysis, designed the study, collected data, and reviewed and revised the manuscript. XT: conceptualized and designed the study, coordinated and supervised data collection, and critically reviewed the manuscript for important intellectual content. All authors contributed to the article and approved the submitted version.

Funding

National Natural Science Foundation of China, Youth Science Foundation Project (grant no. 82202061).

Acknowledgments

We are grateful to the patients and their family for their collaboration.

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

References

1. Tangye SG, Al-Herz W, Bousfiha A, Chatila T, Cunningham-Rundles C, Etzioni A, et al. Human inborn errors of immunity: 2019 update on the classification from the international union of immunological societies expert committee. *J Clin Immunol.* (2020) 40(1):24–64. doi: 10.1007/s10875-019-00737-x

2. Rodero MP, Crow YJ. Type I interferon-mediated monogenic autoinflammation: the type I interferonopathies, a conceptual overview. *J Exp Med.* (2016) 213 (12):2527–38. doi: 10.1084/jem.20161596

3. He T, Qi Z, Luo S, Xia Y, Li C, Huang Y, et al. Clinical and genetic analysis of COPA syndrome in one case. *J Clin Pediatr.* (2018) 36:787–90. doi: 10.3969/j.issn.1000-3606.2018.10.015

4. Guan Y, Liu H, Tang X, Xu H, Peng Y, Zhou C, et al. Effective sirolimus treatment of 2 COPA syndrome patients. *J Allergy Clin Immunol Pract.* (2021) 9(2):999–1001.e1. doi: 10.1016/j.jaip.2020.10.019

5. Zeng J, Hao J, Zhou W, Zhou Z, Miao H. A novel mutation c.841C>T in COPA syndrome of an 11-year-old boy: a case report and short literature review. *Front Pediatr.* (2021) 9:773112. doi: 10.3389/fped.2021.773112

6. Bader-Meunier B, Bustaffa M, Iskounen T, Carter E, Marsh JA, Baujat G, et al. Rheumatoid factor positive polyarticular juvenile idiopathic arthritis associated with a novel COPA mutation. *Rheumatology.* (2021) 60(5):e171–3. doi: 10.1093/rheumatology/keaa763

7. Psarianos P, Kwan JYY, Dell S, Wee WB, Rey-McIntyre K, Chen H, et al. COPA syndrome (Ala239Pro) presenting with isolated follicular bronchiolitis in early childhood: case report. *J Clin Immunol.* (2021) 41(7):1660–3. doi: 10.1007/s10875-021-01082-8

8. Watkin LB, Jessen B, Wiszniewski W, Vece TJ, Jan M, Sha Y, et al. COPA Mutations impair ER-Golgi transport and cause hereditary autoimmune-mediated lung disease and arthritis. *Nat Genet.* (2015) 47(6):654–60. doi: 10.1038/ng.3279

9. Frémond ML, Crow YJ. STING-mediated lung inflammation and beyond. *J Clin Immunol.* (2021) 41(3):501–14. doi: 10.1007/s10875-021-00974-z

10. Taveira-DaSilva AM, Markello TC, Kleiner DE, Jones AM, Groden C, Macnamara E, et al. Expanding the phenotype of COPA syndrome: a kindred with typical and atypical features. *J Med Genet.* (2019) 56(11):778–82. doi: 10.1136/jmedgenet-2018-105560

11. Estrada K, Whelan CW, Zhao F, Bronson P, Handsaker RE, Sun C, et al. A whole-genome sequence study identifies genetic risk factors for neuromyelitis optica. *Nat Commun.* (2018) 9(1):1929. doi: 10.1038/s41467-018-04332-3

12. Leite MI, Coutinho E, Lana-Peixoto M, Apostolos S, Waters P, Sato D, et al. Myasthenia gravis and neuromyelitis optica spectrum disorder: a multicenter study of 16 patients. *Neurology.* (2012) 78(20):1601–7. doi: 10.1212/WNL.0b013e31825644ff

13. Jarius S, Ruprecht K, Wildemann B, Kuempfel T, Ringelstein M, Geis C, et al. Contrasting disease patterns in seropositive and seronegative neuromyelitis optica: a multicentre study of 175 patients. *J Neuroinflammation.* (2012) 9:14. doi: 10.1186/1742-2094-9-14

14. Williams J, McGlasson S, Irani S, Duffy D, Crow Y, Hunt D, et al. Scottish NMOSD study group, et al. Neuromyelitis optica in patients with increased interferon alpha concentrations. *Lancet Neurol.* (2020) 19(1):31–3. doi: 10.1016/S1474-4422(19)30445-4

15. Volpi S, Tsui J, Mariani M, Pastorino C, Caorsi R, Sacco O, et al. Type I interferon pathway activation in COPA syndrome. *Clin Immunol.* (2018) 187:33–6. doi: 10.1016/j.clim.2017.10.001

16. Frémond M-L, Legendre M, Fayon M, Clement A, Filhol-Blin E, Richard N, et al. Use of ruxolitinib in COPA syndrome manifesting as life-threatening alveolar haemorrhage. *Thorax.* (2020) 75(1):92–5. doi: 10.1136/thoraxjnl-2019-213892

17. Krutzke S, Rietschel C, Horneff G. Baricitinib in therapy of COPA syndrome in a 15-year-old girl. *Eur J Rheumatol.* (2019) 7(Suppl 1):1–4. doi: 10.5152/eurjrheum.2019.18177

frontiers | Frontiers in Pediatrics

TYPE Case Report
PUBLISHED 14 March 2023
DOI 10.3389/fped.2023.1096770

OPEN ACCESS

EDITED BY
Che Kang Lim,
Singapore General Hospital, Singapore

REVIEWED BY
Lucia Dora Notarangelo,
Asst of the Brescia Spedali Civili, Italy
Chun Wang,
Shanghai Jiao Tong University, China

*CORRESPONDENCE
Ying Dou
douying523@aliyun.com

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

RECEIVED 12 November 2022
ACCEPTED 10 February 2023
PUBLISHED 14 March 2023

CITATION
Chen Y, Zhou L, Guan X, Wen X, Yu J and Dou Y (2023) Case report: Gene mutations and clinical characteristics of four patients with osteopetrosis.
Front. Pediatr. 11:1096770.
doi: 10.3389/fped.2023.1096770

Case report: Gene mutations and clinical characteristics of four patients with osteopetrosis

Yu Chen[1,2], Lina Zhou[1], Xianmin Guan[2], Xianhao Wen[2], Jie Yu[2] and Ying Dou[1,2]*

[1]National Clinical Research Center for Child Health and Disorders, Ministry of Education Key Laboratory of Child Development and Disorders, Chongqing Key Laboratory of Child Infection and Immunity, Children's Hospital of Chongqing Medical University, Chongqing, China, [2]Department of Hematological Oncology, Children's Hospital of Chongqing Medical University, Chongqing, China

Osteopetrosis is characterized by increased bone density caused by decreased osteoclasts or dysfunction of their differentiation and absorption properties, usually caused by biallelic variants of the *TCIRG1(OMIM:604592)*and *CLCN7 (OMIM:602727)* genes. Herein, the clinical, biochemical, and radiological manifestations of osteopetrosis in four Chinese children are described. Whole-exome sequencing identified compound heterozygous variants of the *CLCN7* and *TCIRG1* genes in these patients. In Patient 1, two novel variants were identified in *CLCN7:c.880T > G(p.F294V)* and *c.686C > G(p.S229X)*. Patient 2 harbored previously reported a single gene variant *c.643G > A(p.G215R) in CLCN7*. Patient 3 had a novel variant *c.569A > G(p.N190S)* and a novel frameshift variant *c.1113dupG(p.N372fs)* in *CLCN7*. Patient 4 had a frameshift variant *c.43delA(p.K15fs)* and variant *c.C1360T* in *TCIRG1*, resulting in the formation of a premature termination codon *(p.R454X)*, both of which were reported previously. Our results expand the spectrum of identified genetic variation in osteopetrosis and provide a deeper understanding of the relations between genotype and clinical characteristics of this disorder.

KEYWORDS

osteopetrosis, CLCN7, TCIRG1, mutation, variants

1. Introduction

Osteopetrosis encompasses a group of rare metabolic bone diseases characterized by impaired osteoclast activity or development, resulting in high bone mineral density (1). The disease is classified into three major clinical subtypes: autosomal recessive osteopetrosis (ARO), intermediate autosomal osteopetrosis (IAO), and autosomal dominant osteopetrosis (ADO). Patients with ARO, the most severe osteopetrosis type, usually present various fatal manifestations soon after birth and die in infancy or before age three years. Apart from the general manifestations of osteopetrosis, patients with ARO can present with pancytopenia, hepatosplenomegaly, cranial nerve compression, and hydrocephalus (2). The clinical course is often serious and, if left untreated, fatal within the first year of life. Diagnosis is challenging and often delayed or misdiagnosed (1). The spectrum of CLCN7-related osteopetrosis includes infantile malignant CLCN7-related ARO, IAO and ADOII.*CLCN7* (13%–16%) and *TCIRG1* (51%–53%) are the major obligate genes responsible for ARO. *CLCN7* encodes the H(+)/Cl(−) exchange transporter 7, which provides chloride conductance across lysosomes in osteoclasts, ensuring the acidification necessary for cellular function (3). ARO is most often caused by biallelic defects in the gene TCIRG1.*TCIRG1* encodes the

a3 subunit, an essential isoform of the vacuolar ATPase proton pump involved in both acidification of the osteoclast resorption lacuna and secretory lysosome trafficking (4, 5). *CLCN7* plays a synergistic role when hydrogen ions are transported outside the cell by *TCIRG1* (6).

2. Case descriptions

Herein, we report the clinical features and genetic findings of four patients with osteopetrosis.

Patient 1, a boy aged 2 months and 21 days, presented initially with anemia, skin hemorrhage, and hepatosplenomegaly. His physical examination showed: pale complexion and moderately pale lips, facial skin, and earlobes; neck palpable, ~5–6 lymph nodes ~1.0cm × 1.0 cm, axillary and abdominal grooves could be palpated, and several lymph nodes enlarged by ~0.5cm × 0.5 cm. Laboratory findings showed white blood cell (WBC) 18.04×10^9/L, platelets (PLT) 40×10^9/L, hemoglobin (Hb) 80 g/L, lymphocyte percentage 0.52, neutrophil percentage 0.2, and lactate dehydrogenase (LDH) 1429.6 U/x Imaging examination showed bilateral axillary lymph nodes, bilateral cervical and inguinal lymph node lesions, and spleen enlargement. Bone density was increased, and the medullary cavity was unclear; transverse translucent bands were seen in the metaphysis of the long bones of the limbs, and thin periosteal shadows were seen in the double radius and femur.

Patient 2 was a girl aged 2 years and 3 months who was followed up for bilateral femoral and tibial fractures. Laboratory results showed WBC 6.92×10^9/L, PLT 256×10^9/L, red blood cells (RBC) 3.92×10^{12}/L, Hb 10^7 g/L, lymphocyte percentage 0.315, neutrophil percentage 0.602, and LDH 914.5 U/L. Imaging showed that the cranial cavity was deformed and elongated, the anteroposterior diameter was prolonged, the cranial plate was thickened to varying degrees, fractures of both femurs and tibias, and slightly widened anterior space of the right femoral neck. The bones of both hips, femurs, tibia, fibula, and osseous bones of the foot were obviously dense. The liver and spleen lymph nodes showed no obvious abnormalities.,The patient had no renal complication,and her father hasn't noticed any symptoms about osteosclerosis.

Patient 3 was a boy aged 1 year and 10 months. His head mass was identified at 1year and 8 months, with a hard quality and swelling around the frame and uneven bone density detected. Laboratory examination showed: WBC 8.93 × 109/L, PLT 245 × 109/L, RBC 5.82 × 1,012/L, Hb 112 g/L, lymphocyte percentage 0.3, neutrophil percentage 0.62, and LDH 599.1 U/ L. Imaging showed slightly enlarged bilateral neck and axillary lymph nodes, without obvious liver or spleen abnormalities. The bone density of bilateral ribs, cervical thoracic vertebra, clavicle and right humerus was uneven, scattered in multiple patchy high-density areas, the boundary was not clear, the double lung texture increased and thickened, and the boundary was blurred.

Patient 4, a boy aged 1 years and 15 days, presented at the hospital five times prior to diagnosis. The first four were due to pulmonary infections, from which he was discharged after improvement with antibiotics. At the fifth admission he presented mainly with anemia, thrombocytopenia, hepatosplenomegaly, cough, and shortness of breath. Physical examination showed slightly pale complexion and bilateral conjunctiva, hyperemic throat, and a mild inspiratory three-concave sign. Laboratory examination showed: WBC 14.34×10^9/L, PLT 45×10^9/L, RBC 2.78×10^{12}/L, Hb 86 g/L, lymphocyte percentage 0.74, neutrophil percentage 0.14, and LDH 617.1 U/L. Imaging showed: slightly enlarged lymph nodes of the liver and spleen; slightly separated left renal pelvis; generally increased density of the craniofacial bones, pelvis, both ends of the ribs, bilateral humerus, ulna, radius, femur, tibia, and fibula. The boundary between the medullary cavity and the cortical bone was unclear, density of the metaphysis was uneven, and different degrees of periosteal reaction could be seen. Radiographs showed bilaterally increased calcaneus and talar bone density. The long bones of the extremities were increased in density, the medullary cavity was unclear, and the metaphysis was difficult to see but slightly widened and thickened, and the lower segment showed translucent shadows. Neurological examination revealed neuromotor retardation, intellectual disability, and right hearing loss. Bone marrow biopsy showed obvious calcification of trabecular bone, intertrabecular fibrosis, and trilineage visible. Some bone marrow hematopoietic cells were significantly squeezed.

2.1. Clinical and laboratory findings

Routine blood tests in Patients 1 and 4 showed that RBC, PLT, and Hb always remained below normal levels (**Figures 1A–C**). Before the sixth test results, Patients 1and 4 received RBC 0.25 U and PLT 0.75 U transfusion treatments, respectively; each soon thereafter fell below normal levels. These results were consistent with long-term anemia in patients with osteopetrosis. WBC findings in Patient 4 were always higher than normal, possibly associated with recurrent respiratory infections (**Figure 1D**). Clinical and laboratory data of the four patients are presented in **Table 1**. Averaged results of multiple assessments, and their standard deviations, were calculated using GraphPad Prism 8.0.

2.2. X-ray examination and bone marrow biopsy

The diagnosis of osteopetrosis was also based on the skeletal radiographs.The four patients in our study,imaging examinations consisting of computerized tomography (CT) scans and x-rays revealed a general increase of bone density involving the skull, vertebrae and limbs. And marrow biopsy results in patient3 showed partial calcification of trabecular bone (**Figure 2**).

2.3. Whole-exome sequencing

Genetic analysis is pivotal to the ARO diagnostic work-up. Genomic DNA samples were extracted from peripheral blood

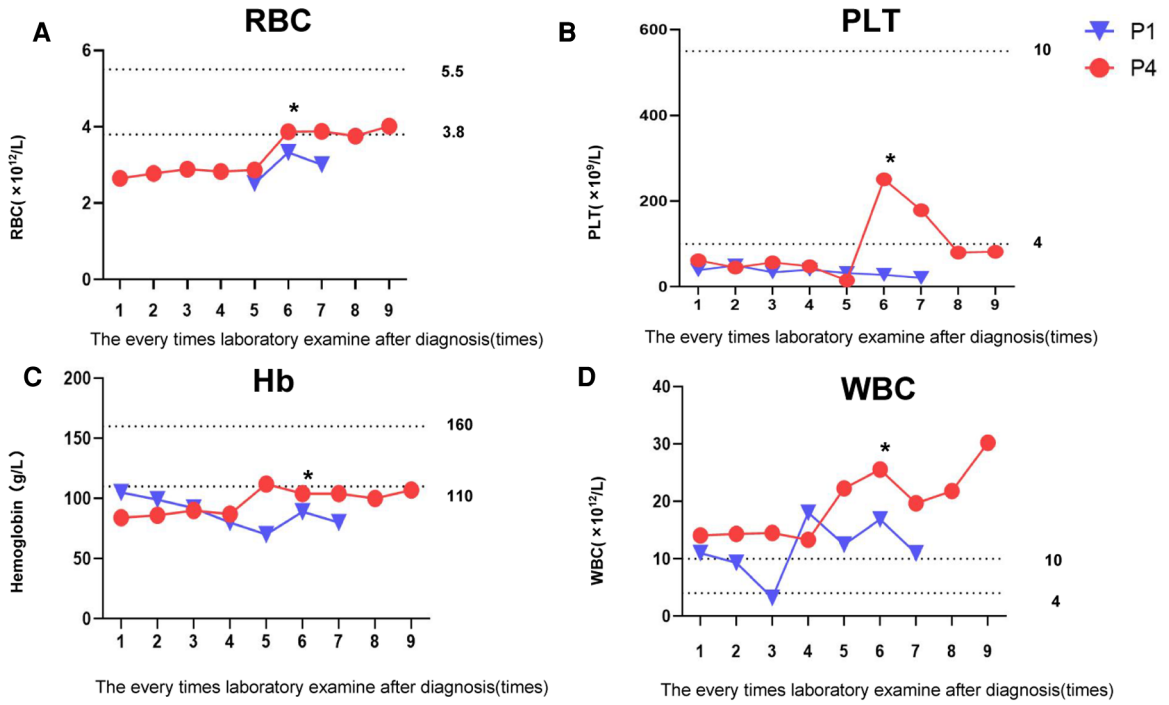


FIGURE 1
Changes in patients' routine blood examination. (A) Changes in patients' RBC levels. (B) Changes in patients' PLT levels. (C) Changes in patients' Hb levels. (D) Changes in patients' WBC levels. *Transfusion treatments: Patient 1 received RBC 0.25 U; Patient 4 received PLT 0.75 U.

using a QIAamp DNA Mini Kit (Qiagen, China) and samples from both the patients and their parents were sent to a third-party testing company (Beijing Hyster Clinical Laboratory). In this method, the target regions of disease-related genes were captured and deeply sequenced with an average sequencing depth of 500–1000x. The overall coverage can reach more than 99%.

The identified variants in *TCIRG1* and *CLCN7* likely caused the manifestations described above.

In Patient 1, two novel variants were identified in the *CLCN7* (NM_001114331) gene: *c.880T > G(p.F294V)* and *c.686C > G (p.S229X)*. Patient 2 harbored a previously reported hotspot variant *c.643G > A(p.G215R)* in *CLCN7* (7). Patient 3 had a novel variant *c.569A > G(p.N190S)* and a novel frameshift variant *c.1113dupG(p.N372fs)* in *CLCN7*. Patient 4 had a frameshift variant *c.43delA(p.K15Fs)* and a variant *c.C1360T* in *TCIRG1*, resulting in the formation of a premature termination codon *(p.R454X)*, both of which have been previously reported (8). And each patient the variants are inherited from their parents (**Figure 3**).

2.4. Pathogenicity predictions of the *c.880t > G(p.F294v)* and *c.569a > G(p.N190s)* variants in the *CLCN7* gene

Potential pathogenicity of the validated missense variant was predicted using SIFT (https://www.siftwallet.com/), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/), MutationTaster (https://www.mutationtaster.org/), MutationAssessor (http://mutationassessor.org/r3/), FATHMM (http://fathmm.biocompute.org.uk/inherited.html), PROVEAN (https://www.jcvi.org/research/provean), and CADD (https://cadd.gs.washington.edu/). NCBI gene (https://www.ncbi.nlm.nih.gov/gene) was used to analyze the conservation of the mutated amino acid. gnomAD(https://www.gnomad-sg.org).

The *CLCN7 p.F294V* and *p.N190S* variants, observed in Patients 1 and 3, respectively, occur in the voltage-gated chloride channel domain, which is highly conserved in multiple species. In addition, amino acids at positions 294 and 190 in the *CLCN7* protein are highly conserved.

Next, we evaluated the pathogenicity of the *c.880T > G (p.F294V)* and *c.569A > G(p.N190S)* variants using seven prediction software types. The prediction results for *c.880T > G (p.F294V)* are listed in **Table 2**, and indicated that this variant is possibly deleterious and implied that it exerts a medium effect on the function of the *CLCN7* protein.And the frequency in general people is 0 which means it's probably pathogenic. The prediction results for *c.569A > G(p.N190S)* are also in **Table 2**, and suggest that this variant is damaging and show that it exerts a low effect on the function of the *CLCN7* protein.And the frequency in general people is $8.179*10^{-6}$. These findings indicate that both variants in the *CLCN7* gene are pathogenic.

3. Discussion

Herein, we reported on four patients with osteopetrosis, in whom we identified compound heterozygous variants in the *CLCN7* and *TCIRG1* genes. Among the variants, *c.880T > G*, *c.686C > G*, *c.569A > G*, and *c.1113dupG* in *CLCN7* are novel.

10.3389/fped.2023.1096770

TABLE 1 Summary of clinical findings of the patients.

Patients	P1	P2	P3	P4
General information				
Sex	Male	Female	Male	Male
Age at diagnosis	2 months 21 days	2 years 3 months	1 year 10 months	1 month 15 days
Initial diagnosis results of blood chemistry				
RBC ($3.8–5.5 \times 10^{12}$/L)	2.95 ± 0.41	3.92	5.82	3.28 ± 0.57
WBC($4–10 \times 10^{9}$/L)	11.7 ± 4.94	6.92	8.93	19.5 ± 5.96
Platelets ($100–550 \times 10^{9}$/L)	34.86 ± 9.32	256	245	91 ± 75.23
MCV (80–100 fL)	87.33 ± 0.80	89.3	58.9	88.25 ± 6.34
MCH (26–32 pg)	27.03 ± 0.67	27.3	19.2	29.82 ± 4.24
MCHC (320–360 g/L)	309.33 ± 5.51	305	327	339.5 ± 54.27
RDW% (<15.5%)	20.33 ± 0.23	18	16.3	19.92 ± 3.03
lymphocytes absolute value ($2–6.5 \times 10^{9}$)	6.4 ± 1.7	2.18	2.68	11.04 ± 3.0
neutrophils absolute value ($1.3–6.7 \times 10^{9}$)	3.82 ± 1.12	4.17	5.54	5.15 ± 2.60
Hemoglobin (110–160 g/L)	87.85 ± 12.10	70	112	97.11 ± 10.43
Total bilirubin (1.8–21 umol/L)	33.4	244.9	2.8	19.23 ± 10.10
Direct bilirubin (0–6.7 umol/L)	13.1	244.9	2.8	15.5 ± 11.90
Albumin (38–52 g/L)	37.9	38.3	41.8	39.4 ± 7.34
ALT (0-40 U/L)	58.3	29.1	26.2	38.7 ± 7.64
AST(0-45 U/L)	194.6	64.6	99.4	53.1 ± 14.14
AST/ALT (0.23–2.47)	3.34	2.22	3.79	1.36 ± 0.99
LDH (110–330U/L)	1429.6	914.5	599.1	662.05 ± 63.57
ALP (<500 IU/L)	546.1	119.4	210.7	451.7
Calcium (2.2–3.0 mmol/L)	2.42	1.88	2.52	2.22 ± 0.04
Phosphorus (1.29–2.26 mmol/L)	0.98	2.54	1.84	1.35 ± 0.247
PT	0.89	11.1		12.15 ± 0.50
APTT (9–12.8)	10.7	12.2		29.8 ± 8.06
Fibrinogen (1.1–3.3 g/L)	1.73	2.02		1.30 ± 0.15
TT (19.5–35.4S)	24.7	18.4		19.15 ± 2.33
D-Dimer (<0.55)	9.3	8.92		21.11 ± 3.42
Other findings				
Hepatosplenomegaly	Yes	No	No	Yes
Increased bone density	Yes	Yes	Yes	Yes

WBC, white blood cell; RBC, red blood cell; RDW%, Red blood cell volume distribution width; PLT, Platelets; Hb, Hemoglobin; LDH, lactate dehydrogenase; ALP, alkaline phosphatase; PT, Prothrombin Time; APTT, activated partial thromboplastin time; TT, Thrombin time.

Two novel missense *CLCN7* variants, *c.880T > G(p.F294V)* and *c.569A > G(p.N190S)*, occurred at positions highly conserved across multiple species. In addition, both variants were predicted to be deleterious by multiple in silico tools (**Table 2**), indicating impaired function of the *CLCN7* protein.

CLCN7 gene mutations are involved in the pathogenesis of various forms of osteopetrosis. It is located on human chromosome 16p13.3, contains 25 exons, and encodes the 803 amino acid chloride channel protein 7 (CIC-7) (3). Chloride channels are responsible for chloride ion transport and play a key role in the generation and transmission of cellular electrical signals. *CLCN7* is a member of the voltage-gated chloride channel protein family, which mediates the exchange of chloride ions for protons and maintains an acidic environment for bone resorption. *CLCN7* is essential for efficient proton pumping, due to its role in neutralizing currents. It is also involved in the secretion of acid into the resorption void, a specialized acidic compartment for mineral bone matrix degradation. Disruption of *CLCN7* expression results in severe lysosomal storage disorders that, in addition to osteopetrosis, can lead to neurodegeneration, including retinal atrophy (9–11).

The *TCIRG1* gene is located on human chromosome 11q13.2 and contains 22 exons (12), encoding the 830 amino acid a3 subunit of V-ATPase. V-ATPase is a proton pump, the main function of which is pumping hydrogen ions into secretory lysosomes. Osteoclasts degrade bone through acidification by vesicle-like V-ATPase (13). When hydrogen ions are pumped out of osteoclasts, they acidify the cortical environment between osteoclasts and bone tissue, promoting bone resorption and regulating bone formation and development. *CLCN7* cooperates with the gene product of the a3 subunit *TCIRG1* of V-ATPase, acting synergistically when hydrogen ions are transported outside the cell by *TCIRG1* (14, 15).

Besides radiology and bone marrow pathology findings are also very important to the differential diagnosis of osteopetrosis helping to distinguish anemia from leukemia.The main pathological disease change is a defect in osteoclast function, which normally degrade bone to initiate remodeling by maintaining formation of calcification groups. Though many osteoclasts remain, they are dysfunctional and thus during osteogenesis the cartilage matrix continues and the calcified osteoclast cannot break down and

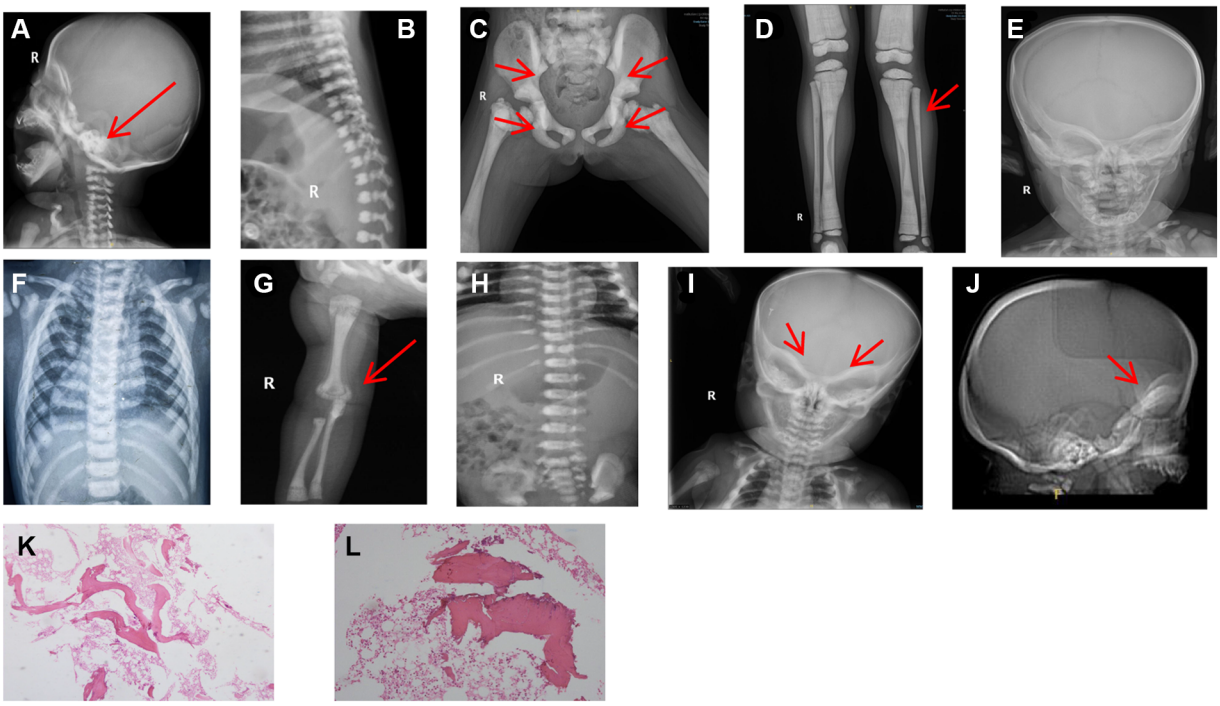


FIGURE 2
Patient x-rays. (A) Patient 1: significant increase in bone density observed at skull base (arrow). (B) Patient 1: vertebral endplate showed typical "sandwich vertebrae" appearance. (C) Patient 2: generalized increase in bone density and typical pelvic wing "bone-in-bone" evident (arrow); high-bone density under the cartilage in the pelvis (arrow). (D) Patient 2: diffusely increased bone density confirmed and osteopetrosis in the distal metaphysis of the femur showed typical "bone-in-bone" appearance (arrow). (E) Patient 3: head mass was identified with a hard quality and swelling around the frame and uneven bone density detected. (F) Patient 3: vertebral endplate showed typical "sandwich vertebrae" appearance. (G) Patient 4: diffuse increase in bone density at the upper radius (arrow), showing "bone-in-bone" appearance. (H) Patient 4: vertebral endplate appeared thickened, with "sandwich vertebrae". (I) Patient 4: observed osteopetrosis in skull base (arrow). (J) Patient 4: lateral skull computerized tomographyshowing increased thickness of skull base (arrow). (K-L) Patient 3: marrow biopsy results showed partial calcification of trabecular bone; HE staining × 40 (K); HE staining × 100 (L).

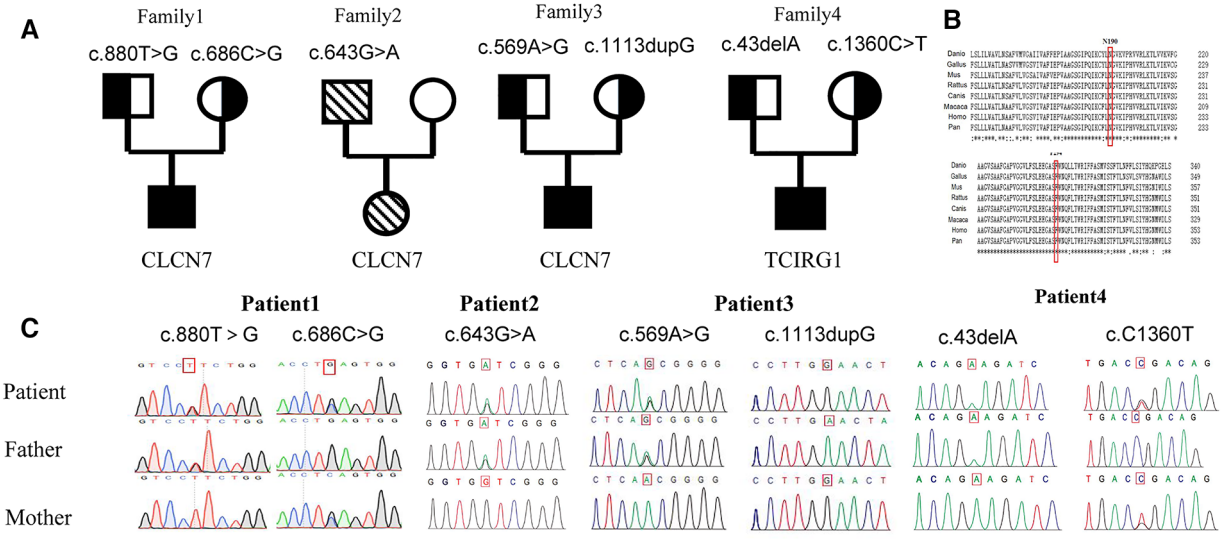


FIGURE 3
Family pedigrees and genetic sequencings. (A) Pedigree for each patient. (B) Genetic analysis of the gene mutation in family of Patient 1: compound heterozygous variants c. 880T > C(*p*. F294V) and c. 686C > G(*p*. S229X) in the CLCN7 gene. Homologous sequences of *CLCN7* across species (C). A cross-species alignment of amino acid sequences showed that *p*. F294V and *p*. N190S variants occur in a highly conserved region.

TABLE 2 Pathogenicity predictions of c. T880G, *p*. F294V and c.569A > G, *p*. N190S in CLCN7 gene.

Prediction software	c.T880G,*p*.F294V		c.569A > G, *p*. N190S	
	Score	Prediction	Score	Prediction
SIFT_score	0.001	Damaging	0	Damaging
Polyphen2	1	Probably_damaging	0.988	Probably_damaging
MutationTaster	1	Disease_causing	1	Disease_causing
MutationAssessor	2.74	Medium (functional impact)	1.47	Low (functional impact)
FATHMM	−3.33	Damaging	−1.75	Damaging
PROVEAN	−6.76	Damaging	−4.83	Damaging
CADD	25.3	Damaging	23.6	Damaging

absorb normal bone resorption activity which calcification of the cartilage and bone tissue cannot be replaced by normal bone tissue and accumulation that bone tissue cannot rebuild bone density. The hard bones brittle fracture easily and marrow cavities shrink, disappear, or occlude (16).

Three major clinical forms of osteopetrosis have been described,: ADO, autosomal recessive osteopetrosis (ARO), and IAO. ADO is the most common and can be roughly divided into three types: ADOI, ADo-II and ADO-III. The expression of ADOI type was relatively mild, mainly caused by *LRP5* gene mutation. *CLCN7* gene is considered to be the main pathogenic gene causing type II, with penetrance ranging from 60% to 90%. Studies on ADOIII type are still lacking, which is mainly caused by *PLEKHM1* gene mutation. Patients 1 and 4 had severe ARO, onset of which usually occurs in the first year of life and is characterized by generalized increased bone density and heterogeneous symptoms, including hematological and neural defects with diverse severity (17). Laboratory results showed that these two patients also had severe anemia, and although both were treated with blood products, they continued to fall below normal levels shortly thereafter.

According to Whyte's research, elevated serum total LDH and AST can distinguish ADO among the sclerosing bone. However, all four patients herein had elevated AST and LDH of varying severities. Mutations of *CLCN7* and *TCIRG1* both impact osteoclasts so that they fail to properly acidify the surrounding cells and the cortical environment between bone tissues, which affects the extra-skeletal tissues and leads to multiple tissue disorders, and ultimately to elevated AST and LDH. The specific mechanisms remain to be further studied in patients with ADO and mouse models (18, 19). At present, genetic diagnosis is still the most important method for confirming osteopetrosis classification.

Patient 2 was diagnosed with ADOII, the main characteristics of this type which are diffuse, including symmetrical osteopetrosis mainly affecting the spine, pelvis, and skull. Severe symptoms include fracture, osteomyelitis, vision loss, and bone marrow failure (20). Among these clinical symptoms, fracture, especially of the long bones, is the most likely independent complication of ADOII. ADOII is an autosomal dominant disorder with incomplete penetrance. The penetrance rate has previously been reported to be between 75% and 94%.There are also asymptomatic carriers of the *G215R* mutation and nonpenetrance rates were 24%. So far, the father of patient 2 has not found any symptoms related to stone bone disease, so he may be an asymptomatic carrier. Almost every adult with ADO (98%) reports fracture. In children with ADO, fractures are less prevalent but still occur in 53% (21). These symptoms were present in Patient 2, who had fractures of both femoral necks. This may be due to impairment of bone remodeling from defective osteoclast function, unsuccessful repair of microdamage, and abnormal bone biomechanics (22). Thus, patients with ADO should be appropriately counseled regarding their propensity for fractures and other ADO-related problems.

Childhood-onset IAO can have autosomal dominant or autosomal recessive inheritance. Its clinical manifestations are intermediate between those of ADO and ARO.The onset of symptoms is usually before the age of first 2 years. Children may present with fractures after minor trauma or characteristic changes on x-rays obtained for other clinical indications. Hematologic signs are milder than those in ARO and are usually restricted to anemia. Inheritance is AD or AR (23). In terms of clinical symptoms, Patient 3 may have this type. Although the clinical symptoms of ADO and IAO are less severe than those of ARO, some patients have been diagnosed with significant clinical manifestations in infancy, which should receive appropriate clinical attention.

Currently, to improve patient survival and quality of life, hematopoietic stem cell transplantation (HSCT) is the only curative treatment for osteopathy. HSCT does not improve the course of osteopetrosis in individuals with *TNFSFII* gene mutation, nor can it reverse the renal tubular acidosis and renal injury caused by *CAII* gene mutation. However, for individuals with *TCIRG1* gene mutation, the effects of transplantation are clearer (1, 24–27). Therefore, genetic testing has guiding significance for whether children can be cured by transplantation. When successful, bone marrow biopsy shows normal results ~13 weeks post-HSCT. In a mouse model of *TCIRG1* defect, bone performance is partly correct on x-ray and histopathological analysis 8 weeks post-HSCT, and after 18 weeks is almost entirely normalized.

To date, Patient 2 has survived without specific treatment, but remains at risk for fractures. Patients 1 and 3 were lost to follow-up. Only Patient 4 met the conditions for HSCT, but his parents were not satisfied with the match result (HLA8/10) and refused treatment. Treatment efficacy depends on patient genotype. Individuals with *OSTM1* gene abnormalities and some *CLCN7* gene mutations associated with neurodegeneration do not show improvement with HSCT (28).

In summary, four children with osteopetrosis were classified based on clinical manifestations, biochemical examinations, radiological changes, and genetic defects. Accurate osteopathy

classification is essential to diagnosis and differentiation from other diseases, as well as to subsequent treatment. Herein we identified novel variants, expanding spectrum of the *TCRIG1* and *CLCN7* genes and deepening our understanding of the relations between genotype and clinical characteristics of osteopetrosis.

Data availability statement

The datasets presented in this article are not readily available because of ethical and privacy restrictions. Requests to access the datasets should be directed to the corresponding author/s.

Ethics statement

The studies involving human participants were reviewed and approved by the Medical Ethics Committee of the Children's Hospital of Chongqing Medical University (approval number: 445/2022). 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 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

Conceptualization: DY and YJ designed the study; YC and LZ wrote the manuscript and analyzed patients'data; XG and XW followed the patients; all authors reviewed the manuscript before publication. All authors contributed to the article and approved the submitted version.

Funding

1. Chongqing Medical University Future Medical Youth Innovation Team Development Support Program (Grant No. W0132). 2. Chongqing medical scientific research project (Joint project of Chongqing Health Commission and Science and Technology Bureau) (Grant No. 2021MSXM112). 3. The National Natural Science Foundation of China Youth Science Fund Project (Grant No. 81601753)

Acknowledgments

We thank the patients and their families for their kind cooperation in this study. We thank the doctors and nurses for supporting 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

References

1. Tolar J, Teitelbaum SL, Orchard PJ. Osteopetrosis. *N Engl J Med*. (2004) 351 (27):2839–49. doi: 10.1056/NEJMra040952

2. Sobacchi C, Schulz A, Coxon FP, Villa A, Helfrich MH. Osteopetrosis: genetics, treatment and new insights into osteoclast function. *Nat Rev Endocrinol*. (2013) 9 (9):522–36. doi: 10.1038/nrendo.2013.137

3. Kornak U, Kasper D, Bösl MR, Kaiser E, Schweizer M, Schulz A, et al. Loss of the ClC-7 chloride channel leads to osteopetrosis in mice and man. *Cell*. (2001) 104 (2):205–15. doi: 10.1016/S0092-8674(01)00206-9

4. Capo V, Abinun M, Villa A. Osteoclast rich osteopetrosis due to defects in the TCIRG1 gene. *Bone*. (2022) 165:116519. doi: 10.1016/j.bone.2022.116519

5. Taranta A, Migliaccio S, Recchia I, Caniglia M, Luciani M, De Rossi G, et al. Genotype-phenotype relationship in human ATP6i-dependent autosomal recessive osteopetrosis. *Am J Pathol*. (2003) 162(1):57–68. doi: 10.1016/S0002-9440(10)63798-4

6. Yu T, Yu Y, Wang J, Yin L, Zhou Y, Ying D, et al. Identification of TCIRG1 and CLCN7 gene mutations in a patient with autosomal recessive osteopetrosis. *Mol Med Rep*. (2014) 9(4):1191–6. doi: 10.3892/mmr.2014.1955

7. Piret SE, Gorvin CM, Trinh A, Taylor J, Lise S, Taylor JC, et al. Autosomal dominant osteopetrosis associated with renal tubular acidosis is due to a CLCN7 mutation. *Am J Med Genet A*. (2016) 170(11):2988–92. doi: 10.1002/ajmg.a.37755

8. Liang H, Li N, Yao RE, Yu T, Ding L, Chen J, et al. Clinical and molecular characterization of five Chinese patients with autosomal recessive osteopetrosis. *Mol Genet Genomic Med*. (2021) 9(11):e1815. doi: 10.1002/mgg3.1815

9. Zifarelli G. The role of the lysosomal cl(-)/H(+) antiporter ClC-7 in osteopetrosis and neurodegeneration. *Cells*. (2022) 11(3):366. doi: 10.3390/cells11030366

10. Brandt S, Jentsch TJ. ClC-6 and ClC-7 are two novel broadly expressed members of the CLC chloride channel family. *FEBS Lett*. (1995) 377(1):15–20. doi: 10.1016/0014-5793(95)01298-2

11. Zifarelli G, Pusch M. CLC Chloride channels and transporters: a biophysical and physiological perspective. *Rev Physiol Biochem Pharmacol*. (2007) 158:23–76. doi: 10.1007/112_2006_0605

12. Koçak G, Güzel BN, Mıhçı E, Küpesiz OA, Yalçın K, Manguoğlu AE. TCIRG1 And SNX10 gene mutations in the patients with autosomal recessive osteopetrosis. *Gene*. (2019) 702:83–8. doi: 10.1016/j.gene.2019.02.088

13. Flinck M, Hagelund S, Gorbatenko A, Severin M, Pedraz-Cuesta E, Novak I, et al. The vacuolar H(+) ATPase α3 subunit negatively regulates migration and invasion of human pancreatic ductal adenocarcinoma cells. *Cells*. (2020) 9(2):465. doi: 10.3390/cells9020465

14. Zhang XY, He JW, Fu WZ, Wang C, Zhang ZL. Novel mutations of TCIRG1 cause a malignant and mild phenotype of autosomal recessive osteopetrosis (ARO) in four Chinese families. *Acta Pharmacol Sin*. (2017) 38(11):1456–65. doi: 10.1038/aps.2017.108

15. Frattini A, Orchard PJ, Sobacchi C, Giliani S, Abinun M, Mattsson JP, et al. Defects in TCIRG1 subunit of the vacuolar proton pump are responsible for a subset of human autosomal recessive osteopetrosis. *Nat Genet*. (2000) 25(3):343–6. doi: 10.1038/77131

16. Quarello P, Forni M, Barberis L, Defilippi C, Campagnoli MF, Silvestro L, et al. Severe malignant osteopetrosis caused by a GL gene mutation. *J Bone Miner Res*. (2004) 19(7):1194–9. doi: 10.1359/JBMR.040407

17. Zhang Z, Xia W, He J, Zhang Z, Ke Y, Yue H, et al. Exome sequencing identifies SLCO2A1 mutations as a cause of primary hypertrophic osteoarthropathy. *Am J Hum Genet*. (2012) 90(1):125–32. doi: 10.1016/j.ajhg.2011.11.019

18. Whyte MP, Kempa LG, McAlister WH, Zhang F, Mumm S, Wenkert D. Elevated serum lactate dehydrogenase isoenzymes and aspartate transaminase distinguish Albers-Schönberg disease (chloride channel 7 deficiency osteopetrosis) among the sclerosing bone disorders. *J Bone Miner Res*. (2010) 25(11):2515–26. doi: 10.1002/jbmr.130

19. Whyte MP, Chines A, Silva DP Jr., Landt Y, Ladenson JH. Creatine kinase brain isoenzyme (BB-CK) presence in serum distinguishes osteopetroses among the sclerosing bone disorders. *J Bone Miner Res*. (1996) 11(10):1438–43. doi: 10.1002/jbmr.5650111010

20. Wang C, Zhang H, He JW, Gu JM, Hu WW, Hu YQ, et al. The virulence gene and clinical phenotypes of osteopetrosis in the Chinese population: six novel mutations of the CLCN7 gene in twelve osteopetrosis families. *J Bone Miner Metab*. (2012) 30(3):338–48. doi: 10.1007/s00774-011-0319-z

21. Waguespack SG, Hui SL, Dimeglio LA, Econs MJ. Autosomal dominant osteopetrosis: clinical severity and natural history of 94 subjects with a chloride channel 7 gene mutation. *J Clin Endocrinol Metab*. (2007) 92(3):771–8. doi: 10.1210/jc.2006-1986

22. Mashiba T, Hirano T, Turner CH, Forwood MR, Johnston CC, Burr DB. Suppressed bone turnover by bisphosphonates increases microdamage accumulation and reduces some biomechanical properties in dog rib. *J Bone Miner Res*. (2000) 15(4):613–20. doi: 10.1359/jbmr.2000.15.4.613

23. Sobacchi C, Villa A, Schulz A, Kornak U, Adam MP, Everman DB, et al. CLCN7-Related Osteopetrosis. In: MP Adam, DB Everman, GM Mirzaa, RA Pagon, SE Wallace, LJH Bean, editors. *Genereviews(®)*. Seattle, WA: University of Washington, Seattle Copyright © 1993-2023, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved (1993).

24. Fasth A, Porras O. Human malignant osteopetrosis: pathophysiology, management and the role of bone marrow transplantation. *Pediatr Transplant*. (1999) 3(Suppl 1):102–7. doi: 10.1034/j.1399-3046.1999.00063.x

25. Corbacioglu S, Hönig M, Lahr G, Stöhr S, Berry G, Friedrich W, et al. Stem cell transplantation in children with infantile osteopetrosis is associated with a high incidence of VOD, which could be prevented with defibrotide. *Bone Marrow Transplant*. (2006) 38(8):547–53. doi: 10.1038/sj.bmt.1705485

26. Johansson M, Jansson L, Ehinger M, Fasth A, Karlsson S, Richter J. Neonatal hematopoietic stem cell transplantation cures oc/oc mice from osteopetrosis. *Exp Hematol*. (2006) 34(2):242–9. doi: 10.1016/j.exphem.2005.11.010

27. Steward CG, Blair A, Moppett J, Clarke E, Virgo P, Lankester A, et al. High peripheral blood progenitor cell counts enable autologous backup before stem cell transplantation for malignant infantile osteopetrosis. *Biol Blood Marrow Transplant*. (2005) 11(2):115–21. doi: 10.1016/j.bbmt.2004.11.001

28. Johansson MK, de Vries TJ, Schoenmaker T, Ehinger M, Brun AC, Fasth A, et al. Hematopoietic stem cell-targeted neonatal gene therapy reverses lethally progressive osteopetrosis in oc/oc mice. *Blood*. (2007) 109(12):5178–85. doi: 10.1182/blood-2006-12-061382

frontiers | Frontiers in Genetics

TYPE Case Report
PUBLISHED 16 March 2023
DOI 10.3389/fgene.2023.1115027

OPEN ACCESS

EDITED BY
Che Kang Lim,
Singapore General Hospital, Singapore

REVIEWED BY
Mario Ernesto Cruz-Munoz,
Universidad Autónoma del Estado de Morelos, Mexico
Sinem Firtina,
University of Istinye, Türkiye

*CORRESPONDENCE
Wei Shi,
shiwei@csu.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 03 December 2022
ACCEPTED 27 February 2023
PUBLISHED 16 March 2023

CITATION
Deng S, Rao S, Wang AR and Shi W (2023), Case report: Rubella virus-induced cutaneous granulomas in a girl with atypical SCID caused by *DCLRE1C* gene mutations.
Front. Genet. 14:1115027.
doi: 10.3389/fgene.2023.1115027

Case report: Rubella virus-induced cutaneous granulomas in a girl with atypical SCID caused by *DCLRE1C* gene mutations

Sihan Deng[1†], Shijia Rao[2†], Alun R. Wang[3] and Wei Shi[1,4*]

[1]Department of Dermatology, Xiangya Hospital, Central South University, Changsha, Hunan, China, [2]Department of Dermatology, The Second Xiangya Hospital, Central South University, Hunan Key Laboratory of Medical Epigenomics, Changsha, Hunan, China, [3]Department of Pathology, Tulane University School of Medicine, New Orleans, LA, United States, [4]National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, Hunan, China

Here, we report a case of rubella virus-induced granulomatous dermatitis in a young girl with immunodeficiency caused by *DCLRE1C* gene mutations. The patient was a 6-year-old girl who presented with multiple erythematous plaques on the face and limbs. Biopsies of the lesions revealed tuberculoid necrotizing granulomas. No pathogens could be identified on extensive special stains, tissue cultures, or PCR-based microbiology assays. Metagenomic next-generation sequencing analysis revealed the rubella virus. Underlying atypical severe combined immunodeficiency was recognized based on the patient's history of repetitive infections since birth, low T-cell, B-cell, and NK cell counts, and abnormal immunoglobulins and complements. Whole-exome sequencing revealed the genetic abnormality of the atypical severe combined immunodeficiency (SCID), and compound heterozygous mutations of the *DCLRE1C* gene were detected. This report highlights the diagnostic values of metagenomic next-generation sequencing in identifying rare pathogens causing cutaneous granulomas in patients with atypical SCID.

KEYWORDS

next-generation sequencing, rubella virus, whole-exome sequencing, *DCLRE1C* gene, granulomatous, severe combined immunodeficiency syndrome

1 Introduction

Cutaneous granulomatous dermatitis is a common disorder that can be divided into infectious and non-infectious etiologies. Infection should be ruled out if no clear cause is identified, especially in immunocompromised patients. Even if an infectious etiology is suspected, the causal microorganisms are often unidentifiable. The prevalence of granulomas is rather high among patients with combined immune deficiency and other types of immunodeficiencies. Skin represents a primary site of granulomas in patients with combined immune deficiency. Some granulomas are caused by the rubella virus infection. The traditional ways of identifying microorganisms are based on morphology, histochemistry, and cultures with limited sensitivity and efficiency. Targeted PCR assays developed in recent years have improved sensitivity significantly. Metagenomic next-generation sequencing (mNGS) is an unbiased assay that identifies a broad spectrum of

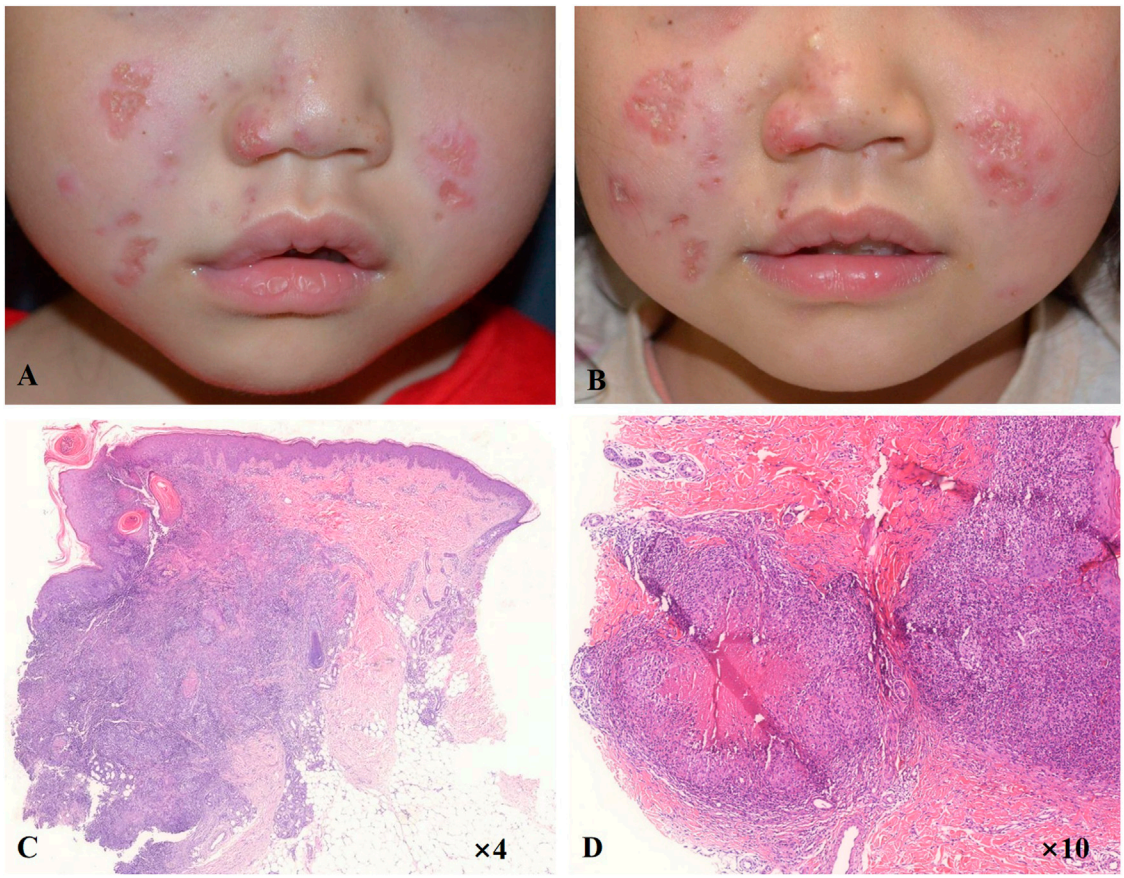


FIGURE 1
RuV infection-induced cutaneous granulomatous dermatitis. Erythematous plaques on cheeks at visit **(A)** and after 10-month follow-up **(B)**. **(C)** Dense lymphohistiocytic infiltrate in the dermis forming a nodule. **(D)** Necrotizing granulomas in the deep dermis.

pathogens, including many rare pathogens. Mutations in the *DCLRE1C* gene have been documented to cause immunodeficiency with phenotypes ranging from severe combined immunodeficiency to mere antibody deficiency. The whole-exome sequencing technique is an effective tool in revealing underlying molecular mechanisms causing immunodeficiencies.

2 Manuscript

2.1 Case description

A 6-year-old girl presented with facial erythematous plaques that had gradually spread into her limbs over the course of a year (Figure 1A). The center of some erythematous plaques was ulcerated and scabbed spontaneously. She appeared to be predisposed to respiratory infections since birth. Multiple biopsies of the plaques revealed necrotizing granulomatous inflammation (Figure 1D). No microorganisms were identified on PAS and acid-fast stains, tissue cultures, and PCR assays for fungi and mycobacteria. Dense lymphocyte infiltrate with lymphoid follicle formation in the dermis with adnexal involvement was also found in a biopsy from the left upper arm (Figure 1C). Most lymphocytes were T-cells ($CD3^+$) with mixed $CD4^+$ and $CD8^+$ cells. Some lymphocytes showed mild cytological atypia. Clonal expansion of the T-cell population was detected through T-cell receptor gene rearrangement studies. The patient had neutropenia and lymphopenia of T cells, B cells, and NK cells. Specifically, cell counts of $CD3^+$ T cells, $CD4^+$ T cells, and $CD8^+$ T cells were all lower than normal. Subpopulations of the CD3+/CD8-/IL-4+ lymphocytes and CD3+/CD8-/IFN-γ+ lymphocytes were higher than normal, and CD3+/CD4+/CD25+/FOXP3+ lymphocytes and CD3+/CD8-/IL-17A + lymphocytes were lower than normal. Serum IgA and IgM levels were significantly lower, and the patient's C3 level was slightly lower than normal (Table 1). No clinically significant nodular metabolic enhancement was observed on PET-CT. The timeline of the patient's care is summarized in Figure 2.

Because mycobacterial infections were not ruled out, a diagnostic therapeutic 8-week course of quadruple antituberculosis therapy was performed but without effect. Repeated metagenomic DNA and RNA mNGS analyses were performed, and rubella virus (RuV, read 7280) was detected. A serology test showed RuV IgM was 1.41 S/CO (positive range:≥1), and the RuV IgG titer was 243.7 IU/mL (positive range:≥10). Therefore, a diagnosis of rubella virus-induced cutaneous granulomatous dermatitis was reached.

TABLE 1 Partial laboratory results of this patient.

	Result	Normal range
$CD3^+$ lymphocytes (%)	**34.93**	50–84
$CD3^+CD4^+$ lymphocytes (%)	51.11	30.00–67.00
$CD3^+CD8^+$ lymphocytes (%)	41.19	23.00–50.00
$CD3^+CD8^-$ lymphocytes (%)	58.81	50.00–75.00
$CD3^+$CD8-IL-4+ lymphocytes (%)	**3.86**	0.30–2.20
$CD3^+$CD8-IFN-γ lymphocytes+ (%)	**38.06**	6.50–28.00
$CD3^+CD4^+$CD25+FOXP3+ lymphocytes (%)	**3.25**	4.10–9.40
$CD3^+$CD8-IL-17A + lymphocytes (%)	0.65	0.20–2.40
$CD4^+$ cell (/uL)	**258**	550–1,440
$CD8^+$ cell (/uL)	**185**	320–1,250
B lymphocyte (%)	**3.62**	5–18
B lymphocyte (/uL)	**23**	90–560
NK cell (/uL)	**128**	150–1,100
Total T lymphocyte (/uL)	**486**	955–2,860
C3 (g/L)	**0.6**	0.7–1.4
C4 (g/L)	0.215	0.1–0.4
IgG (g/L)	5.59	5.4–13.4
IgM (g/L)	**0.11**	0.43–1.52
IgA (g/L)	**0.12**	0.3–1.48
IgE (IU/mL)	1	0–100
Blood routine		
WBC (×10⁹/L)	**2.52**	3.5–9.5
Neutrophils (×10⁹/L)	**1.44**	1.8–6.3
Lymphocytes (×10⁹/L)	**0.49**	1.1–3.2
Monocytes (%)	**12.7**	3–10

In bold: values above or below the reference range.

Considering RuV-associated granulomas are rare and almost exclusively occur in the context of immune deficiency, and the patient had signs of immunodeficiency, a whole- exome sequencing was performed, and heterozygous mutations in *DCLRE1C* gene as c.352G>T (p.G118X) and c.328C>G (p.L110V) were detected (Figure 3). It was concluded that hypomorphic heterozygous mutations of *DCLRE1C* genes led to an atypical SCID in this girl, who developed RuV-induced cutaneous granulomatous dermatitis, likely after inoculation with the rubella virus through vaccination. The patient had been waiting for hematopoietic stem-cell transplantation (HSCT) for the previous 10 months, and no other treatments had been initiated. An increase in the number and size of plaques on her face and limbs had been noted (Figure 1B), and the plaques had remained asymptomatic. The patient was hospitalized for pre-transplant examination and preparation.

2.2 Discussion

Here, we report the first Chinese patient with RuV-induced cutaneous tuberculoid necrotizing granulomas in an immunodeficient child with heterozygous *DCLRE1C* gene mutations. Our patient received an MMR vaccine injection when she was 18 months old. The lesions were likely induced by vaccine-derived live attenuated rubella viruses. In addition, a clonal expansion of the T-cell population in the granulomatous lesions was detected. No obvious evidence of a cutaneous T-cell lymphoma was identified. However, it has been observed that immunodeficiency could lead to a clonal expansion of lymphocyte populations, potentially as an autoreactive proliferation of T cells in response to infections or autoantigens, as likely happened in this patient (Herber et al., 2020).

In 2014, vaccine-derived rubella virus was identified from cutaneous granulomas in patients with primary immunodeficiency (Bodemer et al., 2014). Since then, no more than 100 cases have been reported worldwide, mostly in European and American populations. Pathologically, RuV infections could present as necrotizing granulomas, although they can also present as non-specific dermal chronic inflammation (Buchbinder et al., 2019). Skin lesions have been documented to develop several years on average after measles–mumps–rubella (MMR) vaccination, for which attenuated rubella virus is administered. Wild-type RuV has also been detected much less often in granulomatous lesions (Shields et al., 2021).

Based on the reported cases, all RuV-associated granuloma patients present with some degree of immune deficiency, although some are clinically immunocompetent adults (Wanat et al., 2022). Many patients

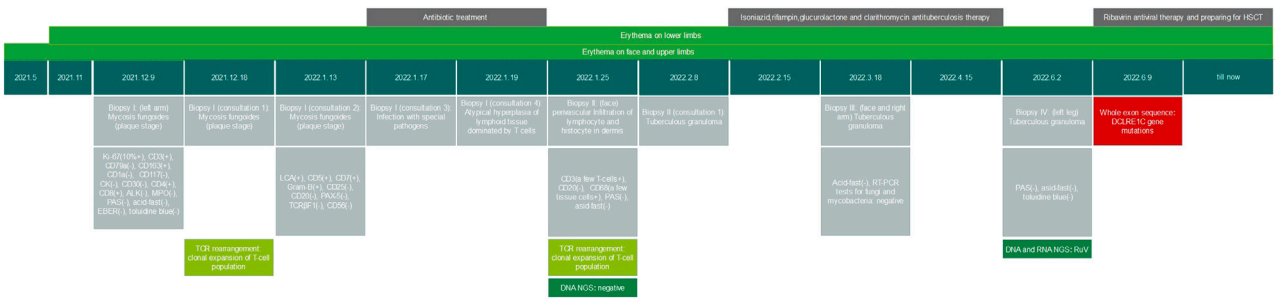


FIGURE 2
Timeline of the patient's care.

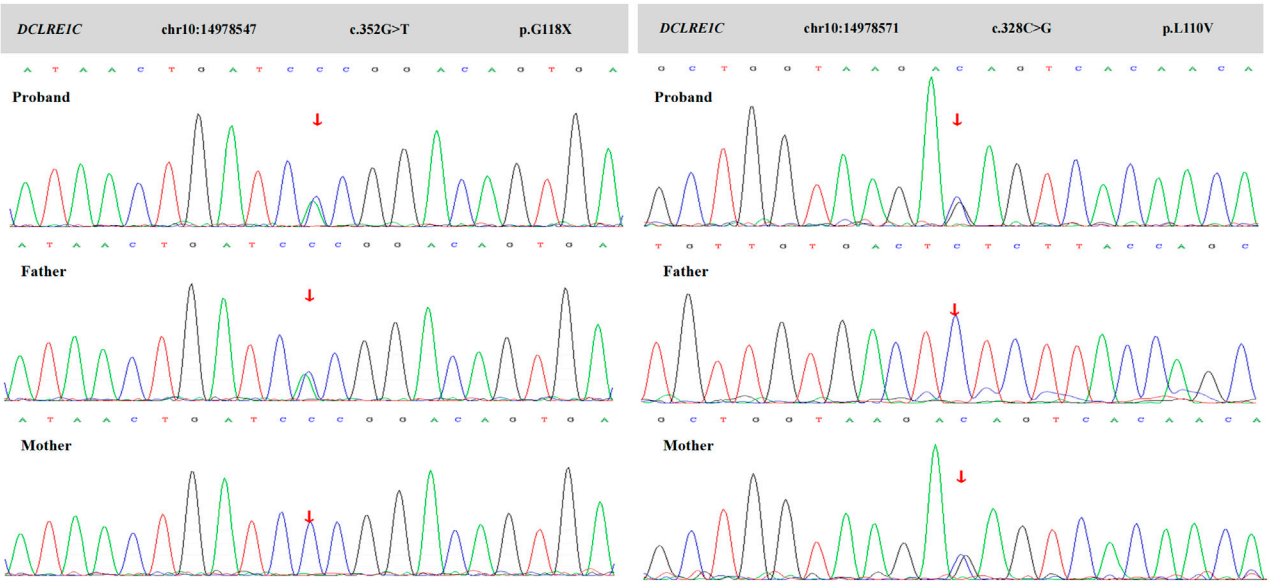


FIGURE 3
Sanger sequence of this family. Chromatograms show that her father is heterozygous for c.352G>T (p.G118X), and her mother is heterozygous for c.328C>G (p.L110V). The proband shows a compound heterozygote of these mutations.

have DNA repair defects, with *DCLRE1C* gene mutations as one of the causes. *DCLRE1C* encodes ARTEMIS, which is essential in the V(D)J recombination of the immunoglobulin and T-cell receptor genes in T- and B-cell development, as well as in DNA repair (Felgentreff et al., 2015). Mutations in the *DCLRE1C* gene can lead to severe combined immunodeficiency, Omenn syndrome, and radiosensitivity (Strubbe et al., 2021). Severe combined immunodeficiency is an inherited, most severe form of primary immunodeficiency caused by mutations in genes involved in lymphocyte development and function and characterized by the absence or dysfunction of T lymphocytes. It affects both cellular and humoral adaptive immunity. B lymphocytes and NK cells may be affected as well. Over 20 different molecular defects are documented. These include defects in genes involved in antigen receptor gene rearrangement, T-cell receptor signaling, T-cell differentiation, thymic development, and thymic egress of T cells.

Patients with hypomorphic mutations, however, may present with less severe clinical phenotypes depending on residual activity levels on the alleles, causing so-called leaky or atypical SCID (Felgentreff et al., 2015). Atypical SCID occurs and may be defined as a primary immunodeficiency disease secondary to hypomorphic mutations in SCID-causing genes with a milder presentation and higher T-cell counts than typical SCID. The T-cell count is generally above 100/μL (Felgentreff et al., 2011). In atypical SCID, patients usually survive beyond 12 months of age. Atypical SCID has been diagnosed in all age groups, frequently significantly later than SCID.

Our patient had a long history of repetitive respiratory tract infections and recurrent nasal stuffiness since birth. These infections are generally not severe and could be alleviated after a few days on antibiotics. A high copy number of Epstein–Barr virus (EBV) DNA was detected in her serum, which would be an indicator that the patient was experiencing an active EBV infection. The patient presented with reduced numbers of peripheral T cells, B cells, NK cells, and neutrophils, and low IgM and IgA levels. Despite these abnormalities, she seemed relatively normal in development and overall health. Her conditions, therefore, fit into atypical SCID.

The pathogenesis of overall white blood cells and immunoglobulin abnormalities caused by *DCLRE1C* gene mutations remains unknown. Usually, the *DCLRE1C* gene variants cause $T^-/B^-/NK^+$ SCID. However, our patient presents with a unique $T^-/B^-/NK^-$ phenotype. She presents with a novel compound heterozygosity of two mutations, c.352G>T and c.328C>G. Mutation of c.352G>T (p.G118X) is a nonsense mutation that would very likely result in gene dysfunction. No mutation of this site has been reported in the database, and the ClinVar database assesses it as pathogenic. The clinical significance of mutation of c.328C>G (p.L110V) is uncertain according to ACMG guidelines. Mutations at this location have been reported, although clinical phenotypes (Shahbazi et al., 2019; Xiao et al., 2021). The presence of compound mutations could make a clear correlation between particular genotypes and phenotypes difficult or impossible.

There are no effective conventional therapies for RuV-associated granulomatous diseases. The efficacy of antiviral therapy, TNF-α, IL-1R antagonists, and glucocorticoids remains uncertain (Perelygina et al., 2021). HSCT is considered a curative treatment, and the treatment is more likely to succeed when it is performed early in the disease process. The caveat is that patients with Artemis-deficient immunodeficiency have a decreased tolerance to the alkylating agents used as preparative regimens for HSCT and that carry a risk of significant long-term toxicity (Schuetz et al., 2014). Therefore, it is important to explain the benefits and risks of HSCT in detail to the patients or their parents. Hematopoietic stem cell transplantation is underway for our patient.

Metagenomic next-generation sequencing is a powerful new platform that can simultaneously identify genetic material from entirely different kingdoms of organisms. It can sequence all nucleic acids in a sample and identify multiple populations of

microorganisms from different taxa and their proportions. The potential clinical applications are tremendous, particularly in the diagnosis of infectious diseases. As a less biased, sensitive, and broad-spectrum assay, it opens doors for revealing pathogens in numerous unsolvable cases per traditional methodologies.

Data availability statement

The data sets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ncbi.nlm.nih.gov/, PRJNA905930.

Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

WS was responsible for planning this work. SD and SR were responsible for drafting this manuscript. WS and AW were responsible for revising this manuscript. WS was responsible for the final approval of the article.

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.2023.1115027/full#supplementary-material

References

Bodemer, C., Sauvage, V., Mahlaoui, N., Cheval, J., Couderc, T., Leclerc-Mercier, S., et al. (2014). Live rubella virus vaccine long-term persistence as an antigenic trigger of cutaneous granulomas in patients with primary immunodeficiency. *Clin. Microbiol. Infect.* 20 (10), O656–O663. doi:10.1111/1469-0691.12573

Buchbinder, D., Hauck, F., Albert, M. H., Rack, A., Bakhtiar, S., Shcherbina, A., et al. (2019). Rubella virus-associated cutaneous granulomatous disease: A unique complication in immune-deficient patients, not limited to DNA repair disorders. *J. Clin. Immunol.* 39 (1), 81–89. doi:10.1007/s10875-018-0581-0

Felgentreff, K., Lee, Y. N., Frugoni, F., Du, L., van der Burg, M., Giliani, S., et al. (2015). Functional analysis of naturally occurring DCLRE1C mutations and correlation with the clinical phenotype of ARTEMIS deficiency. *J. Allergy Clin. Immunol.* 136 (1), 140–150. doi:10.1016/j.jaci.2015.03.005

Felgentreff, K., Perez-Becker, R., Speckmann, C., Schwarz, K., Kalwak, K., Markelj, G., et al. (2011). Clinical and immunological manifestations of patients with atypical severe combined immunodeficiency. *Clin. Immunol.* 141 (1), 73–82. doi:10.1016/j.clim.2011.05.007

Herber, M., Mertz, P., Dieudonné, Y., Guffroy, B., Jung, S., Gies, V., et al. (2020). Primary immunodeficiencies and lymphoma: A systematic review of literature. *Leuk. Lymphoma* 61 (2), 274–284. doi:10.1080/10428194.2019.1672056

Perelygina, L., Faisthalab, R., Abernathy, E., Chen, M. H., Hao, L., Bercovitch, L., et al. (2021). Rubella virus infected macrophages and neutrophils define patterns of granulomatous inflammation in inborn and acquired errors of immunity. *Front. Immunol.* 12, 796065. doi:10.3389/fimmu.2021.796065

Schuetz, C., Neven, B., Dvorak, C. C., Leroy, S., Ege, M. J., Pannicke, U., et al. (2014). SCID patients with ARTEMIS vs RAG deficiencies following HCT: Increased risk of late toxicity in ARTEMIS-deficient SCID. *Blood* 123 (2), 281–289. doi:10.1182/blood-2013-01-476432

Shahbazi, Z., Yazdani, R., Shahkarami, S., Shahbazi, S., Hamid, M., Sadeghi-Shabestari, M., et al. (2019). Genetic mutations and immunological features of severe combined immunodeficiency patients in Iran. *Immunol. Lett.* 216, 70–78. doi:10.1016/j.imlet.2019.10.001

Shields, B. E., Perelygina, L., Samimi, S., Haun, P., Leung, T., Abernathy, E., et al. (2021). Granulomatous dermatitis associated with rubella virus infection in an adult with immunodeficiency. *JAMA Dermatol* 157 (7), 842–847. doi:10.1001/jamadermatol.2021.1577

Strubbe, S., De Bruyne, M., Pannicke, U., Beyls, E., Vandekerckhove, B., Leclercq, G., et al. (2021). A novel non-coding variant in DCLRE1C results in deregulated splicing and induces SCID through the generation of a truncated ARTEMIS protein that fails to support V(D)J recombination and DNA damage repair. *Front. Immunol.* 12, 674226. doi:10.3389/fimmu.2021.674226

Wanat, K. A., Perelygina, L., Chen, M. H., Hao, L., Abernathy, E., Bender, N. R., et al. (2022). Association of persistent rubella virus with idiopathic skin granulomas in clinically immunocompetent adults. *JAMA Dermatol* 158 (6), 626–633. doi:10.1001/jamadermatol.2022.0828

Xiao, F., Lu, Y., Wu, B., Liu, B., Li, G., Zhang, P., et al. (2021). High-frequency exon deletion of DNA cross-link repair 1C accounting for severe combined immunodeficiency may Be missed by whole-exome sequencing. *Front. Genet.* 12, 677748. doi:10.3389/fgene.2021.677748

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 12 June 2023
DOI 10.3389/fimmu.2023.1160397

OPEN ACCESS

EDITED BY
Jinqiao Sun,
Fudan University, China

REVIEWED BY
Tomas Drgon,
United States Food and Drug Administration, United States
Erin E. Baschal,
University of Colorado Anschutz Medical Campus, United States

*CORRESPONDENCE
Zu-Guo Yu
yuzg@xtu.edu.cn

RECEIVED 09 February 2023
ACCEPTED 29 May 2023
PUBLISHED 12 June 2023

CITATION
Wen Y-P and Yu Z-G (2023) Identifying shared genetic loci and common risk genes of rheumatoid arthritis associated with three autoimmune diseases based on large-scale cross-trait genome-wide association studies.
Front. Immunol. 14:1160397.
doi: 10.3389/fimmu.2023.1160397

Identifying shared genetic loci and common risk genes of rheumatoid arthritis associated with three autoimmune diseases based on large-scale cross-trait genome-wide association studies

Ya-Ping Wen[1,2] and Zu-Guo Yu[1,2]*

[1]National Center for Applied Mathematics in Hunan, Xiangtan University, Hunan, China, [2]Key Laboratory of Intelligent Computing and Information Processing of Ministry of Education, Xiangtan University, Hunan, China

Introduction: Substantial links between autoimmune diseases have been shown by an increasing number of studies, and one hypothesis for this comorbidity is that there is a common genetic cause.

Methods: In this paper, a large-scale cross-trait Genome-wide Association Studies (GWAS) was conducted to investigate the genetic overlap among rheumatoid arthritis, multiple sclerosis, inflammatory bowel disease and type 1 diabetes.

Results and discussion: Through the local genetic correlation analysis, 2 regions with locally significant genetic associations between rheumatoid arthritis and multiple sclerosis, and 4 regions with locally significant genetic associations between rheumatoid arthritis and type 1 diabetes were discovered. By cross-trait meta-analysis, 58 independent loci associated with rheumatoid arthritis and multiple sclerosis, 86 independent loci associated with rheumatoid arthritis and inflammatory bowel disease, and 107 independent loci associated with rheumatoid arthritis and type 1 diabetes were identified with genome-wide significance. In addition, 82 common risk genes were found through genetic identification. Based on gene set enrichment analysis, it was found that shared genes are enriched in exposed dermal system, calf, musculoskeletal, subcutaneous fat, thyroid and other tissues, and are also significantly enriched in 35 biological pathways. To verify the association between diseases, Mendelian randomized analysis was performed, which shows possible causal associations between rheumatoid arthritis and multiple sclerosis, and between rheumatoid arthritis and type 1 diabetes. The common genetic structure of rheumatoid arthritis, multiple sclerosis, inflammatory bowel disease and type 1 diabetes was explored by these studies, and it is believed that this important discovery will lead to new ideas for clinical treatment.

KEYWORDS

rheumatoid arthritis, autoimmune diseases, association studies, shared genes, cross-trait

1 Introduction

It is well known that the major function of the immune system is to protect the host from environmental agents such as microbes or chemicals, thereby preserving the integrity of the body (1). When the body is injured or invaded by pathogenic microorganisms, acute inflammatory reaction is often accompanied, and the immune system and inflammatory mechanism are inseparable (2). However, uncontrolled inflammatory and immune responses can lead to immune system disorders that trigger autoimmune diseases, such as rheumatoid arthritis (RA), inflammatory bowel disease (IBD) and type 1 diabetes mellitus (T1D) (3). Autoimmune diseases are complex diseases caused by genetic and environmental factors (4). The clinical manifestations of these diseases are familial clustering, and multiple immune diseases can occur simultaneously in the same individual, which indicates that autoimmune diseases have a common genetic background. Moreover, genomic studies have shown that the same gene loci can be found in various autoimmune diseases, and genetic overlap exists in autoimmunity, indicating that autoimmune diseases may have the same molecular mechanism (5, 6).

RA is a chronic, inflammatory autoimmune disease that can cause severe movement impairment and deterioration of quality of life (7). In twin and familial studies, the overall heritability of rheumatoid arthritis is estimated to be about 50%-65% (8). It's suggested that rheumatoid arthritis is familial, individuals with a family history of rheumatoid arthritis are at increased risk of developing rheumatoid arthritis due to common genetic factors (9–11). Multiple sclerosis (MS) is an inflammatory autoimmune disease in which the myelin sheath and spinal cord in the central nervous system are damaged, which can result in demyelination and axonal loss. Some studies have suggested that patients with multiple sclerosis have an increased risk of rheumatoid arthritis (12–14). IBD is a chronic non-specific inflammatory condition of the gastrointestinal tract. In recent years, studies have found that the gene predictive risk of RA is positively correlated with the increased risk of IBD (15, 16). Yang et al. studied the common genetic structure of MS and IBD through large-scale genome-wide association studies (GWAS), and the results showed that the comorbidity of MS and IBD has a biological basis (17). In addition, previous studies have shown that individuals with RA, MS or IBD have an increased risk of influenza and related complications (18), and an increased risk of depression, especially in women compared to men (19). Diabetes mellitus (DM) is a chronic disease that causes hyperglycemia due to defective insulin secretion or impaired biological action. DM is a group of physiological dysfunctions characterized by hyperglycemia resulting directly from insulin resistance, inadequate insulin secretion, or excessive glucagon secretion (20). Type 1 diabetes (T1D) is an endocrine disorder in which pancreatic β cells stop producing insulin, typically due to autoimmune destruction (21). Studies have shown that RA is associated with abnormal glucose metabolism, which may lead to the development of DM (22, 23), and patients with MS may increase the risk of developing T1D (24). Ahmad and Ahsan have revealed common risk genes of RA and MS, MS and T1D through reported familial and genetic links (25). Andersen et al. have shown that when parents have RA, IBD, or DM, offspring are at increased risk (26). Recently, Zhao et al. collected summary statistics from GWAS about seven autoimmune diseases, including celiac disease (CEL), MS, primary biliary cirrhosis (PBC), RA, ulcerative colitis (UC), SLE, and T1D to analyze genetic associations (27). Although there is an epidemiological association between RA, MS, IBD and T1D, whether this reflects a common genetic etiology is unclear. Therefore, the purpose of this paper is to reveal the genetic relationships of RA, MS, IBD, and T1D through large-scale cross-trait GWAS analysis.

GWAS combining multiple diseases have become useful tools to identify risk locus for autoimmune diseases, genetic variant associated with multiple diseases, and biological pathways associated with diseases (28–31). Based on the hypothesis that there is a common genetic cause between autoimmune diseases, in this study, we use GWAS summary statistics to investigate the shared genetic capacity of RA, MS, IBD and T1D at the individual variation level and at the genome-wide level, respectively. Firstly, the genetic relationships between RA and MS, RA and IBD, RA and T1D are analyzed. The global genetic associations among diseases are analyzed by linkage disequilibrium score regression (LDSC), and the local genetic associations among diseases are estimated by using ρ−HESS. Then cross-trait meta-analysis is used to identify the shared genetic components between RA and MS, RA and IBD, RA and T1D. Genome-wide association analysis and transcriptome association studies are used to identify potentially the common risk genes among RA, MS, IBD and T1D. Finally, Mendelian randomization is used to analyze the causal relationship between RA and MS, RA and IBD, and RA and T1D respectively. In summary, we leverage large-scale GWAS summary statistics data and preceding genetic methods to gain insight into mechanistic links among RA, MS, IBD and T1D. Our purpose is to identify the common risk genes among RA, MS, IBD and T1D, and provide biological interpretation for common risk genes.

2 Materials and methods

2.1 Datasets

For summary statistics from GWAS about rheumatoid arthritis (RA), multiple sclerosis (MS), inflammatory bowel disease (IBD), the GWAS summary-level data are downloaded from a publicly accessible database GeneATLAS (32). Specifically, the RA meta-analysis summary statistics include 5082 cases and 447182 controls, MS meta-analysis summary statistics include 1406 cases and 450858 controls and IBD meta-analysis summary statistics combine 3878 cases and 448386 controls. The total 452264 samples are all European-ancestry individuals from UK Biobank, and we used 623944 genotype variants that passed quality control. Summary statistics about type 1 diabetes (T1D) (PMID: GCST90013791) which was uploaded on 02/22/2021 (33) were downloaded from the database NHGRI-EBI GWAS Catalog (34). The T1D meta-analysis summary statistics include 181,214 individuals of European ancestry with 6,294,828 genotype variants. The numbers for cases

and controls are not provided in the NHGRI-EBI GWAS Catalog database, but we need not to use this kind of information in our study.

2.2 Methods

2.2.1 LD score regression analysis

To evaluate the genetic correlation between RA and MS, RA and IBD, RA and T1D, the linkage disequilibrium score regression (LDSC) (35) was applied to assess the genetic correlation r_g between RA and MS, IBD, T1D. We applied LDSC to estimate SNP heritability and LD-score intercept for RA, MS, IBD and T1D, respectively. European-ancestry population of 1000 Genome Project Phase 3 (36) was used as reference panel, from which 1.2 million SNPS were obtained for pre-calculated LD-scores.

2.2.2 Local genetic correlations analysis

To investigate whether there are local genetic correlations between RA and MS, RA and IBD, RA and T1D, we estimated the local genetic correlations between each pair of traits in pre-specified LD independent segments using ρ–HESS (37). The LD-independent blocks are used to calculate local heritability and genetic covariance. However, when we calculated the local genetic correlation using 623,944 SNPs (RA, MS, IBD), we found that there are empty loci on chromosome 1 (chr1:178944309-178954470) and chromosome 7 (chr7:124156805-124167552) in 1703 pre-designated independent fragments, so we combined these loci with nearby loci. Accounting for Bonferroni correction, if $P_{\rho-HESS} < 0.05/1701(2.93\times10^{-5})$, there are significant genetic correlations between RA and MS, RA and IBD. For RA and T1D, in addition to the above two regions, there were three empty locus on chr2:95326452-98995201, chr6:29737971-30798168, chr15:20001200-21131604, and we combined these locus with nearby loci, so the significant threshold is $P_{\rho\text{-hess}} < 0.05/1698$ (2.94×10^{-5}).

2.2.3 Cross-trait meta-analysis

After estimating the genetic correlations between RA and MS, RA and IBD, RA and T1D, we used R packet cross phenotype association (CPASSOC) (38) to analyze the GWAS cross-trait association. CPASSOC includes Shet (for heterogeneous data) and Shom (for homogenous data). We applied the PLINK (39) clustering function to identify the independent and significant SNPs, and the parameter is set as –cluster-p1 1.6×10^{-8}–cluster-p2 1×10^{-5} –cluster-r2 0.2 –cluster-kb 500, indicating that SNPs with a P-value less than 1×10^{-5}, r^2 greater than 0.2 and less than 500 kb away from the peak value will be assigned to the cluster of the peak value.

2.2.4 Genome-wide gene-based analysis

In gene-based analysis, genetic variation is annotated, i.e., SNPs correspond to the corresponding gene according to the position on the chromosome, and gene-based association analysis is carried out. The MAGMA (40) analytical model uses multiple linear principal component regression to detect the correlation between genes and the disease. In this study, MAGMA gene analysis is used to identify significant genes associated with RA, MS, IBD and T1D, respectively. Using European-ancestry population of 1000 Genomes Project Phase 3 as reference and Genome Reference Consortium Human Build 37 (hg19) as the SNP locations for gene annotation, we found that 301949 (48.39%) of the total 623944 SNPs are mapped to 17446 genes.

2.2.5 Transcriptome-wide association analysis

To detect genes associated with RA, MS, IBD, and T1D in different tissues, we performed transcriptome-wide association analysis by using e-MAGMA (41). e-MAGMA transforms genome-wide association summary statistics into gene-level statistics by assigning risk variants to its putative genes based on tissue-specific eQTL information. We used eQTL information from 47 tissues of the GTEx (version 8) reference panel (42). In total, we performed TWASs for each trait, one tissue-trait pair at a time.

2.2.6 Enrichment analysis and protein-protein interaction network analysis

In order to understand the underlying biological pathways for the identified shared risk genes in RA with MS, IBD, and T1D, we used the tool Enrichr web server (43) to implement the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. The significant criterion is that the adjusted p-value less than 0.05. In addition, we applied STRING (version 10) (44) to analyze the protein-protein interaction (PPI) network.

2.2.7 Mendelian randomization analysis

We performed MR analysis using MR-PRESSO (45) between RA and MS, RA and IBD, RA and T1D since they are genetically correlated. We built the MR instruments based on LD-independent SNPs.

3 Results

3.1 Genetic correlation between RA and MS, RA and IBD, RA and T1D

We evaluated the global genetic correlation of RA and MS, RA and IBD, RA and T1D using LD score regression (LDSC). RA has the strongest genetic correlation with MS, with a correlation coefficient of 0.4289, followed by RA and IBD, with a correlation coefficient of 0.3743, and then RA and T1D, with a correlation coefficient of -0.3157 (Table 1). Furthermore, the LD score intercepts for RA, MS, IBD and T1D are 0.9982 (Se = 0.0097), 1.0172 (Se = 0.0105), 1.0156 (Se = 0.011) and 0.9933 (Se = 0.0098), respectively, indicating that most of the inflation is due to polygenic effect rather than population structure or sample overlap (46).

Accounting for Bonferroni correction the local genomic regions around individual RA loci from GWAS showed signals of genetic overlap with MS (Figure 1). Although RA and MS have positive global genetic correlation using LDSC, we identified two regions

TABLE 1 Genetic correlation of RA and MS, RA and IBD, RA and T1D.

Trait 1	Trait 2	r_g	$r_g - s_e$	p-value	g_{cov}	$g_{cov} - s_e$
RA	MS	0.4289	0.2932	0.1434	0.0019	0.0012
	IBD	0.3743	0.1504	0.0128	0.0036	0.0013
	T1D	-0.3157	0.0951	0.0009	-0.0102	0.0034

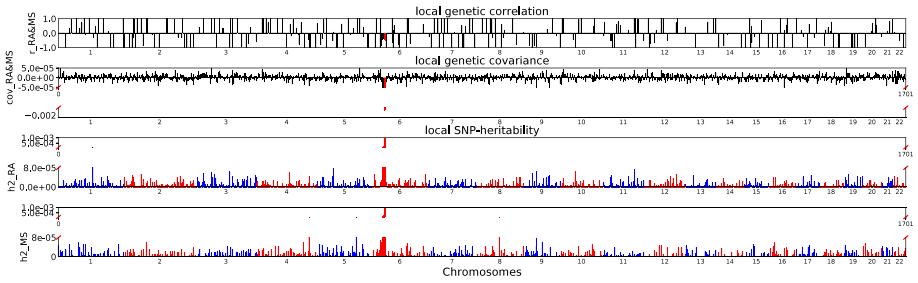


FIGURE 1
Local genetic correlation and local SNP-heritability between RA and MS. For each sub-figure, the top sub-figure represents local genetic correlation, the second represents local genetic covariance. In these two sub-figures, significant local genetic correlation and covariance after multiple testing correction are highlighted in red. Bottom two sub-figures represent local SNP-heritability for individual trait. The X-axis represents the chromosome, h2_RA represents the SNP-heritability of RA, h2_MS represents the SNP-heritability of MS, r_RA-MS represents the genetic correlation between RA and MS, cov_RA-MS represents the genetic covariance between RA and MS.

(chr6:31571218-32682664, $P_{\rho-HESS} = 5.82\times10^{-17}$, and chr6:32682664-33236497, $P_{\rho-HESS} = 2.72 \times 10^{-12}$) that show genome wide significant negative local genetic correlation between RA and MS using heritability estimation from summary statistics (ρ−HESS). This reverse result may be caused by the different definitions of SNP heritability and genetic covariance between ρ−HESS and LDSC. We used ρ−HESS method to evaluate local genetic correlations between RA and IBD. There is no significant local genetic correlated regions (Figure 2), this means that the genetic association between RA and IBD is likely to be shared genetic variants across the genome rather than strong associations in specific genomic regions. The local genetic correlation between RA and T1D is negative in the chromosome 1 and the chromosome 6 regions (Figure 3). There are four significant local genetic correlated regions (chr6: 32682664-33236497, $P_{\rho-HESS} = 2.90 \times 10^{-14}$, chr1:113273306-114873845, $P_{\rho-HESS} = 1.22 \times 10^{-11}$, chr6: 31571218-32682664, $P_{\rho-HESS} = 5.73 \times 10^{-10}$,and chr6: 33236497-35455756, $P_{\rho-HESS} = 1.26 \times 10^{-6}$).

3.2 Identification of risk SNPs from cross-trait meta-analysis of RA, MS, IBD and T1D

We conducted cross-trait meta-analysis to identify risk SNPs that may share association with RA and MS, RA and IBD, RA and T1D using the Cross Phenotype Association (CPASSOC) package (PCPASSOC < 5×10−8/3(1.6×10−8)).

After excluding SNPs that are genome-wide significant in the respective single-trait GWAS, 58 independent loci reached genome-wide significance for RA and MS, 27 of which have been verified to be significantly related to RA and/or MS by GWAS Catalog database verification (Supplementary Table S1). Although most of the independent loci we found were located in the MHC region, we also found loci in the non-MHC region. The loci rs6679677 (on chr1) is close to the PTPN22 gene. The gene involved in regulating CBL function in T cell receptor signaling pathway, and mutations in this gene may be linked to a range of autoimmune diseases rheumatoid arthritis (47). The loci rs7731626 (on chr5) is mapped to ANKRD55 gene which is associated with RA (48) and MS (49).

86 independent loci reached genome-wide significance for RA and IBD, which 46 in this locus have previously been associated to RA and/or IBD (Supplementary Table S2). The loci rs3130695 is mapped to including HLA-B and HLA-C genes which from the HLA class I region is associated with RA (13, 50). The locus rs34213882 and rs9263717 are mapped to HLA-C genes. The genes had genome-wide significant association with IBD. The loci rs11465802 (on chr1), rs11209026 and rs3024505 (on chr1) are mapped to IL23R, C1orf141, and IL10 genes associated with IBD (51). In addition, there are loci rs6679677(on chr1) which is mapped to PTPN22 gene associated with RA. The loci rs1801274(on chr1) is mapped to FCGR2A gene which is associated with RA and IBD. The loci rs2076756 (on chr16) is mapped to NOD2 associated with RA.

107 independent loci reach genome-wide significance for RA and T1D, 47 of which have previously been associated to RA and/or T1D (Supplementary Table S3). Loci rs2856997, rs2070121, rs7383287 are mapped to HLA-DOB genes which are associated with RA. Loci rs2534674, rs2534671, rs6915833 are mapped to MICB genes which are associated with RA (Ancha et al., 2023). Loci rs1150755, rs12198173 are mapped to APOM genes which are associated with RA. Loci rs2233977, rs20547, rs1063646, rs9263719,

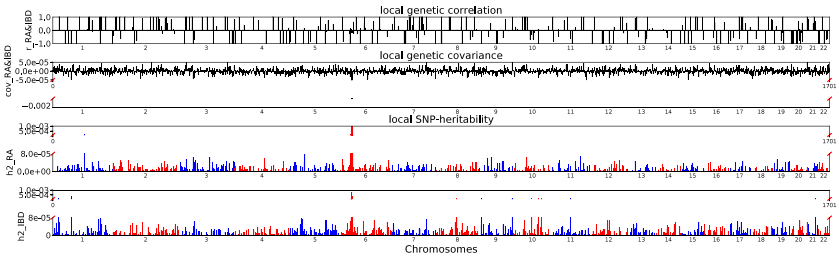


FIGURE 2
Local genetic correlation and local SNP-heritability between RA and IBD. For each sub-figure, the top sub-figure represents local genetic correlation, the second represents local genetic covariance, bottom two sub-figures represent local SNP-heritability for individual trait. The X-axis represents the chromosome, h2_RA represents the SNP-heritability of RA, h2_IBD represents the SNP-heritability of IBD, r_RA-IBD represents the genetic correlation between RA and IBD, cov_RA-IBD represents the genetic covariance between RA and IBD.

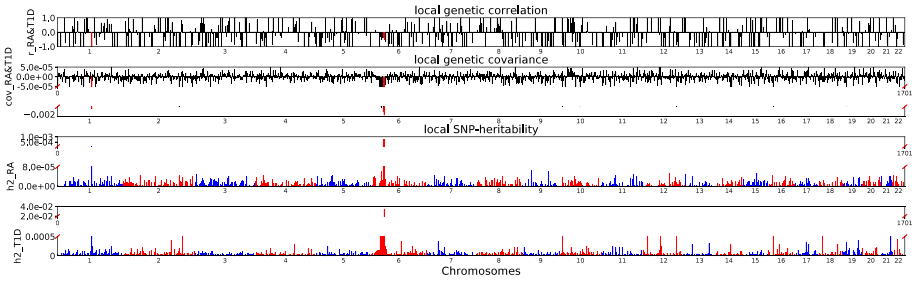


FIGURE 3
Local genetic correlation and local SNP-heritability between RA and T1D. For each sub-figure, the top sub-figure represents local genetic correlation, the second represents local genetic covariance. In these two sub-figures, significant local genetic correlation and covariance after multiple testing correction are highlighted in red. Bottom two sub-figures represent local SNP-heritability for individual trait. The X-axis represents the chromosome, h2_RA represents the SNP-heritability of RA, h2_T1D represents the SNP-heritability of T1D, r_RA-T1D represents the genetic correlation between RA and T1D, cov_RA-T1D represents the genetic covariance between RA and T1D. In the bottom two subfigures, we did not use the same y-axis as Figures 1, 2 because the difference on the order of magnitude for h2_T1D and h2_RA is too large.

rs3094663, rs6916921, rs3819299 and rs12614 are mapped to C6orf15, PSMB9, PSORS1C1, PSORS1C2, NFKBIL1, HLA-B and CFB genes which are associated with RA (52–54). In addition to the above findings in genes located in the MHC region, we also found RA-related loci in non-MHC regions. The loci rs7041847 (on chr9) is mapped to GLIS3, and the loci rs7200786 (on chr16) is mapped to CLEC16A which are associated with RA (55, 56). The loci rs2281808 (on chr20) is mapped to SIRPG, and the rs1805761(on chr12) is mapped to M6PR which are associated with T1D (57, 58). The loci rs6859219 (on chr5) is mapped to ANKRD55 genes which are associated with RA and T1D. Locus rs2847281 and rs7234029 (on chr18) are mapped to PTPN2 genes which are associated with RA and T1D (59, 60). These locus rs705708 (on chr12), rs706778 (on chr10), rs6669008 (on chr1), rs1788103 (on chr18) and rs9976767(on chr21) are mapped to ERBB3, IL2RA, MAGI3, CD226 and UBASH3A which are associated with RA and T1D.

3.3 Genes identified by genome-wide and transcriptome-wide studies

We conducted MAGMA genome-wide gene-based analysis to identify genes associated with RA, MS, IBD, and T1D, respectively. The numbers of genes identified are shown in Figure 4. It can be seen from the figure that after Bonferroni correction of the total 19427 genes, 93 genes ($P_{MAGMA} < 2.87\times10^{-6}$) are identified as significantly correlated with RA; 64 genes are related to MS; 23 genes are associated with IBD; 334 genes are related to T1D (Supplementary Table S4). There are 56 overlapping genes between RA and MS; 10 overlapping genes between RA and IBD; 89 overlapping genes RA and T1D; 10 overlapping genes RA, MS and IBD; 55 overlapping genes RA, MS and T1D; 9 overlapping genes RA, IBD and T1D; 9 among the four diseases (Table 2).

Moreover, we carried out eMAGMA transcriptome-wide gene-based analysis with RA, MS, IBD, and T1D, respectively, and the result are shown in Figure 5. The genes significantly associated with 47 tissues of each disease are identified successively, 147, 140, 174 and 522 genes significantly associated with RA, MS, IBD and T1D are identified, respectively (Supplementary Tables S5-S8). There are 123 overlapping genes between RA and MS; 82 overlapping genes between RA and IBD; 137 overlapping genes RA and T1D; 82 overlapping genes RA, MS and IBD; 122 overlapping genes RA, MS and T1D; 81 overlapping genes RA, IBD and T1D; 81 among the four diseases, eight out of nine common risk genes detected by MAGMA are also detected by e-MAGMA (Supplementary Table S9), we identified 82 common risks among the four diseases. 40 of

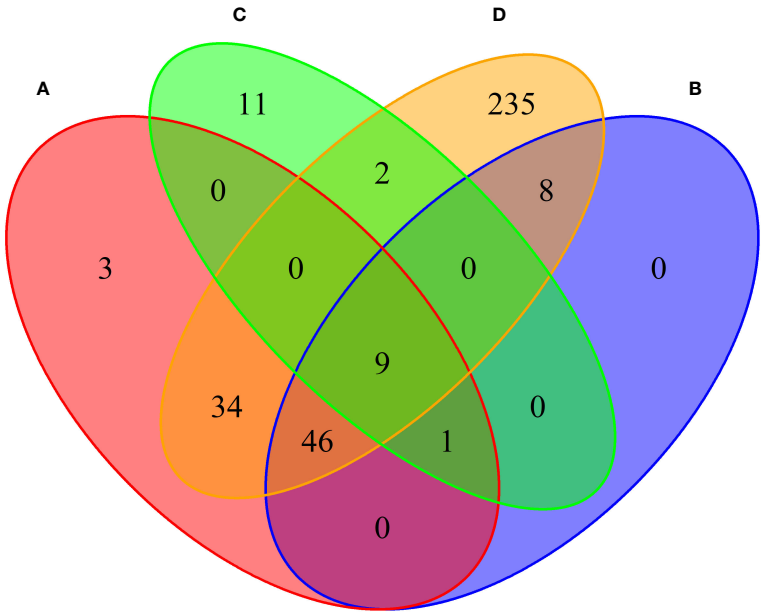


FIGURE 4
Venn diagram of the number of genes identified by MAGMA method. **(A)** RA associated genes, **(B)** MS associated genes, **(C)** IBD associated genes, **(D)** T1D associated genes.

the 82 risk genes are significantly associated with the disease reported in previous studies. We also paid attention to the enrichment analysis of 82 common risk genes in tissues. As shown in Figure 6, it was found that 34 risk genes are mainly enriched in integumentary system of skin sun exposed lower leg, 32 are enriched in muscle skeletal, and 32 are enriched in adipose subcutaneous.

3.4 KEGG enrichment analysis and PPI network analysis results

To understand the impact of risk genes in biological pathways, we used Enrichr to enrich the co-risk genes in a KEGG functional analysis. We found that 35 biological pathways are significantly enriched. As shown in the Table 3, the strongest enrichment signal is antigen processing and presentation, which include 13 enriched genes (HLA-DRB5, HSPA1L, HLA-B, TAP2, HLA-C, TAP1, HLA-A, HLA-F, HLA-G, HLA-E, HLA-DMA, HLA-DOB, HLA-DQA2). In order to understand the interaction between common risk genes of four diseases, we used STRING for PPI network analysis. The 82 risk genes have 279 gene pairs interacting in PPI network, with an average clustering coefficient of 0.421, and the composite scores of all the interacting genes is not less than 0.4, among which the score of 58 gene pair composite score were more than 0.95. The five hub genes (degrees > 15) that extensively interact with other genes in the PPI network are HLA-B, HLA-A, HLA-C, PSMB9, HLA-F. The PPI network of common risk genes are shown in the Figure 7.

TABLE 2 Shared risk genes for RA, MS, IBD and T1D in MAGMA analysis.

Gene	Position	No. SNPs	P_{MAGMA}			
			RA	MS	IBD	T1D
BTNL2	Chr6:32362513-32374900	39	8.77e-49	1.08e-22	4.79e-11	5.06e-90
HLA-DRA	Chr6:32407619-32412823	51	7.42e-39	1.60e-40	3.52e-8	6.01e-82
ATF6B	Chr6:32083045-32096017	16	5.10e-34	1.72e-13	8.85e-8	7.91e-80
EHMT2	Chr6:31847536-31865464	20	1.96e-32	9.18e-24	1.05e-9	5.18e-61
HLA-DQB1	Chr6:32627241-32634466	12	2.89e-32	8.09e-14	5.67e-12	1.09e-44
TAP2	Chr6:32789610-31865464	48	3.03e-22	1.19e-12	1.71e-08	1.63e-121
TRIM31	Chr6:30070674-30080867	31	2.01e-08	1.02e-7	7.90e-10	2.01e-52
NELFE	Chr6:31919864-31926864	14	1.96e-22	1.34e-14	5.19e-10	1.05e-68
MICA	Chr6:31367561-31383090	38	3.11e-13	1.07e-7	1.23e-6	8.61e-145

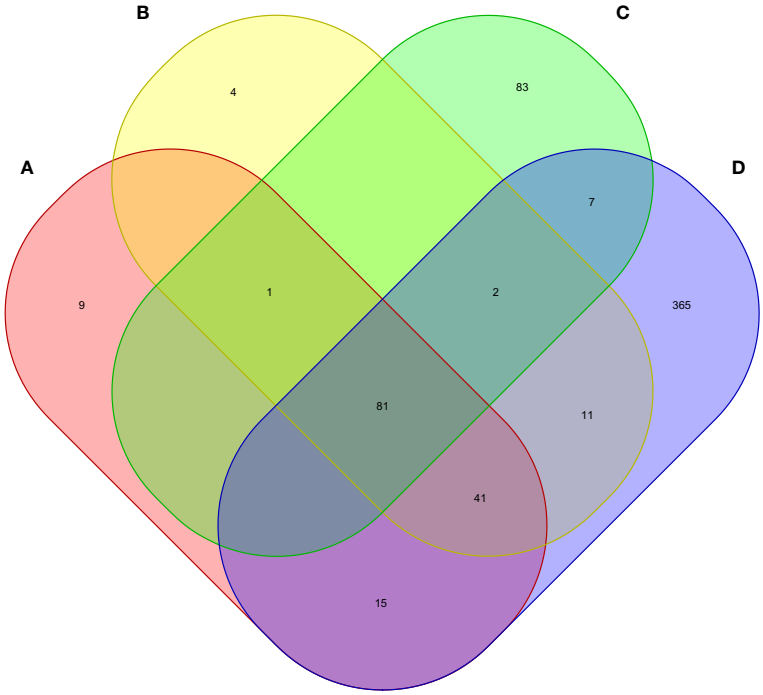


FIGURE 5
Venn diagram of the number of genes identified by e-MAGMA method. **(A)** RA associated genes, **(B)** MS associated genes, **(C)** IBD associated genes, **(D)** T1D associated genes.

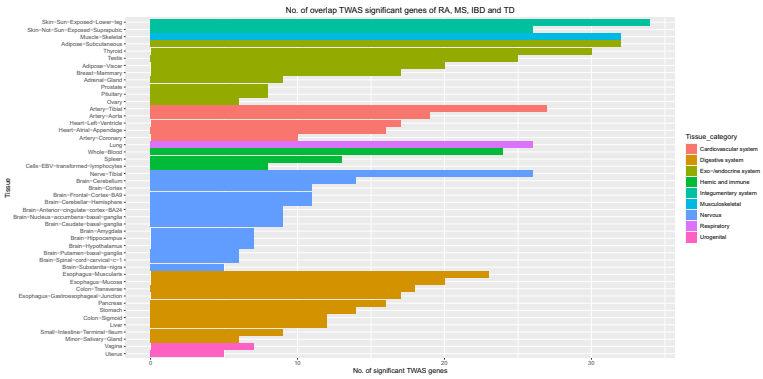


FIGURE 6
Number of overlap TWAS significant genes of RA, MS, IBD and T1D traits across 47 tissues. The X axis shows the count of genes. The Y axis lists tissues. Colors represent different tissue categories. TWAS, transcriptome-wide association studies, No. number.

3.5 Instrumental variable analysis

Finally, we used MR-PRESSO instrumental analysis to develop evidence for causality in the relationship between RA and MS, RA and IBD, RA and T1D, and the results are shown in Table 4. As shown in the Table 4, the finding that there may be a causal relationship between T1D and RA. We found a possible new causal relationship between T1D and RA. Although no relevant studies have confirmed the causal relationship between T1D and RA, the risk of type 1 diabetics developing RA later in life may be attributed in part to the presence of the PTPN22 allele (61).

Another finding is that there may be a causal relationship from MS to RA, but not vice versa, which supports the idea that common immunologic pathways, involving IL-17 and Th17, may be one of the mechanisms through which MS increases susceptibility to RA (17). MS diagnosis increased the likelihood of a patient's subsequent diagnosis of rheumatoid arthritis. Our MR-PRESSO analysis showed no causal relationship between RA and IBD after adjusting pleiotropy. These results further support our findings

TABLE 3 The KEGG pathway was significantly enriched in 82 common risk genes.

Pathway	No. Genes	Adjusted P-value	Pathway	No. Genes	Adjusted P-value
Antigen processing and presentation	13	6.51e-16	Systemic lupus erythematosus	6	1.07e-04
Allograft rejection	10	1.75e-14	Viral carcinogenesis	7	1.07e-04
Graft-versus-host disease	10	3.50e-14	Rheumatoid arthritis	5	2.04e-04
Type I diabetes mellitus	10	3.50e-14	Cellular senescence	6	2.07e-04
Autoimmune thyroid disease	10	2.75e-13	Intestinal immune network for IgA production	4	2.07e-04
Viral myocarditis	10	5.43e-13	Endocytosis	7	3.52e-04
Phagosome	13	7.43e-13	Toxoplasmosis	5	4.17e-04
Epstein-Barr virus infection	12	4.84e-10	Inflammatory bowel disease	4	6.05e-04
Cell adhesion molecules	10	5.85e-9	Leishmaniasis	4	0.00112
Human T-cell leukemia virus 1 infection	11	1.60e-8	Th1 and Th2 cell differentiation	4	0.00213
Herpes simplex virus 1 infection	13	1.10e-6	Hematopoietic cell lineage	4	0.00271
Human cytomegalovirus infection	9	3.15e-06	Th17 cell differentiation	4	0.00351
Natural killer cell mediated cytotoxicity	7	9.16e-06	Influenza	4	0.01894
Kaposi sarcoma-associated herpesvirus infection	8	9.62e-06	Tuberculosis	6	0.02151
Human immunodeficiency virus 1 infection	8	1.71e-05	Longevity regulating pathway	3	0.02763
Staphylococcus aureus infection	6	1.71e-05	Primary immunodeficiency	2	0.03339
Asthma	4	4.78e-05	ABC transporters	2	0.04484
Human papillomavirus infection	9	5.04e-05			

that the shared genetic effects between RA and IBD are more likely to be pleiotropic effects, rather than causal etiology or mechanism.

4 Discussion

In this study, we aimed to determine the genetic relationships among RA, MS, IBD, and T1D by large-scale cross-trait GWAS analysis. Firstly, LDSC is used to identify the genome-wide genetic relationships between RA and MS, RA and IBD, RA and T1D. We found that there are statistically significant genetic relationships between RA and MS, RA and IBD, RA and T1D. Secondly, ρ−HESS is adopted to identify the local genetic relationships between RA and MS, RA and IBD, and RA and T1D. It was found that there are two significant local genetic correlation regions between RA and MS, and four significant local genetic correlation regions between RA and T1D. Thirdly, the CPASSOE method is used to identify significant correlation loci between RA and MS, RA and IBD, RA and T1D. It was found that there are 58 significant correlation loci between RA and MS, 86 significant correlation loci between RA and IBD, and 107 significant correlation loci between RA and T1D. Fourthly, by using the multiple omics method MAGMA and e-MAGMA to identify the common risk genes for four diseases, we found that 82 risk genes show significant association with all four diseases, and 40 of these diseases have been confirmed to be associated with at least one disease. Fifth, we introduced the biological functions of the 82 risk genes found through tissue and organ enrichment analysis, biological pathway enrichment analysis and protein-protein analysis, and found that 82 common risks genes are mainly concentrated in skin sun exposed lower leg, muscle skeletal, adipose subcutaneous, and 35 biological pathways. Finally, we used the MR-PRESSO method to identify the causal relationship between RA and MS, RA and IBD, RA and T1D, and found that there may be causal relationship between RA and T1D, RA and MS, but there is no causal relationship between RA and IBD. The reason of the genetic relationship between RA and IBD is due to pleiotropy effects.

In this study, 82 common risk genes related to RA, MS, IBD and T1D were identified, among which a large number of common genes were found in the HLA region, which plays an important role in immune response. Immune response is one of the main factors affecting RA, MS, IBD and T1D (12, 51, 62). TSBP1 gene has been reported to be associated with four diseases (63–66). Although CCHCR1 gene has been reported to be associated with MS, IBD and T1D diseases (67–69), it may also be important for RA. Twenty-two genes (FLOT1, VARS2, POU5F1, MICA, MICB, NFKBIL1, TAP2, TAP1, BRD2, TNXB, AGPAT1, TRIM31, APOM, TRIM27, SLC44A4, RNF39, AGPAT1, ABCF1, RNF5, CYP21A2, PSORS1C1, LST1) have been reported to be associated with at least one disease. Although no relevant study shows the correlation between TRIM26 and RA, MS, IBD, T1D, we found that TRIM26 is a member of the TRIM protein family, encoded in the locus of major histocompatibility complex Class I region, and TRIM26 interacts with TAB1 and specifically catalyzes K11-linked

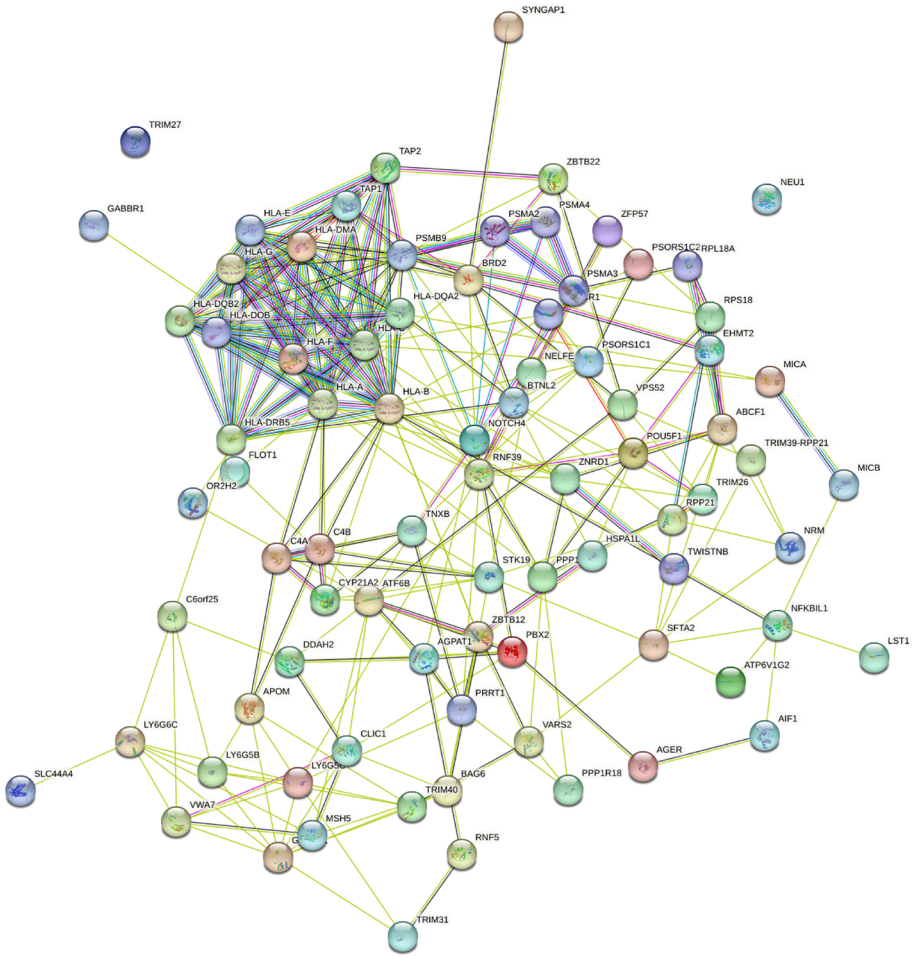


FIGURE 7
The PPI network of 82 common risk genes.

TABLE 4 Mendelian Randomization analysis between RA, MS, IBD and T1D.

Exposure	Outcome	Causal Estimate	Sd	T-stat	P-value[1]
RA	MS	-0.000756071	0.005449628	-0.138738	0.889728
MS	RA	-0.1195902	0.05367141	-2.228192	0.03170483
RA	IBD	-0.006132929	0.003543663	-1.730675	0.08381547
IBD	RA	0.01654898	0.02420687	0.683648	0.4970677
RA	T1D	-0.7789012	0.5110154	-1.524222	0.1276694
T1D	RA	-307.0967	80.94547	-3.793871	0.0002637433

polyubiquitination of TAB1, which facilitates TAK1 activation and initiates downstream signaling, and finally positively regulated TLRS-mediated inflammatory cytokines (70). AIF1 is a 17kDa cytoplasmic calcium-binding inflammatory scaffold protein, which is mainly expressed in immune cells. AIF1 promotes the expression of inflammatory mediators such as cytokines, chemokines and inducible nitric oxide synthase (iNOS), promoting the proliferation and migration of inflammatory cells (71). CLIC1 participates in inflammatory processes by regulating macrophage phagosomal functions such as pH and proteolysis, CLIC1 regulates DC phagosomal pH to ensure optimal processing of antigen for presentation to antigen-specific T-cells (72).

There are also some limitations in this paper. Firstly, Bonferroni correction is the most stringent multiple testing correction method. In genome-wide association analysis, in order to control the probability of false positives, the threshold is often adjusted with

Bonferroni correction. However, due to linkage disequilibrium between GWAS variants, there may be cases where multiple variants or SNPs are linked to each other, so it is not entirely correct to assume that each association test of a GWAS dataset is independent. Therefore, applying the Bonferroni correction usually gives us the most conservative p-value threshold. Because it is too conservative, it often leads to the generation of false negatives, and there may be few variants in the entire genome whose associated p-values can meet this standard. In this study, to investigate whether there are local genetic correlations between RA and MS, RA and IBD, and RA and T1D, we estimated the local genetic correlations between each pair of traits using ρ–HESS (37). The ρ–HESS method used Bonferroni correction as a threshold to identify the local genetic correlation between each pair of traits in the visualization results, so in the subsequent research analysis, we also used Bonferroni correction. Secondly, our results cannot be used to be representative of the global population or children, as the sample of RA, MS and IBD were taken from UK Biobank, individuals of European descent aged between 40 and 69 years, and the T1D summary statistics which individuals are European descent from the NHGRI-EBI GWAS Catalog. Thirdly, due to the lack of biological information at the individual level of genotype and phenotype datasets, we cannot determine whether the effect of polymorphic genes on disease risk is directed. Experimental studies are required to verify the pleiotropic genes identified in this study. Fourthly, existing gene annotation is not comprehensive, which leads to some SNPs not annotating genes. Fifthly, the vast majority of the risk SNPs and shared risk genes we identified are located in the MHC region on chromosome 6, although we used the clustering function of PLINK to identify independent and significant SNPs, due to the extensive linkage disequilibrium in the MHC region, it is possible that the risk SNPs and shared risk genes we identified are correlated. Furthermore, the functions of the newly identified shared risk genes are still unclear, and further studies are needed to understand the functions of the genes and their roles in pathophysiology. The function of the newly discovered shared risk genes is unclear, and there is still a long way to go to understand the function of genes and their role in the pathophysiology of disease. Finally, we did not analyze the combination of genetic and environmental factors that are known to influence autoimmune diseases, including smoking, diet, exercise, and medication. To sum up, further studies are needed to emphasize and explore the biological explanations, and efforts should be made to translate the findings into clinical research or practice. This study provides an effective theoretical basis for future research on the pathogenesis of autoimmune diseases, improvement of diagnostic methods and development of targeted therapies.

5 Conclusion

In this paper, strong genetic associations between RA and three autoimmune diseases have been explored. Through genetic estimation, it was found that there are local genetic correlation signals between RA and MS, RA and T1D. By cross-trait meta-analysis, it was found that there are independent genetic loci related to RA and MS, RA and IBD, RA and T1D. Based on gene correlation analysis, 82 common risk genes were found among the four diseases. Common risk genes are enriched in skin sun exposed lower leg, muscle skeletal, adipose subcutaneous, and 35 biological pathways. Through Mendelian randomization analysis, we found that there may be causal relationship between RA and T1D, RA and MS. Therefore, this study is helpful for the clinical treatment of RA, MS, IBD and T1D.

Data availability statement

Publicly available datasets were analyzed in this study. The data can be found here: http://geneatlas.roslin.ed.ac.uk/ and https://www.ebi.ac.uk/gwas/.

Author contributions

Y-PW contributed to the conception and design of the study, developed the method and wrote the manuscript. Z-GY gave the ideas and supervised the project, also revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by funds from the National Key Research and Development Program of China (grant number: 2020YFC0832405); the National Natural Science Foundation of China (grant number: 11871061); The High-Level Talent Aggregation Project in Hunan Province, China Innovation Team (grant number: 2019RS1060), the Science and Technology Innovation Program of Hunan Province of China (grant number: 2022WK2009) and Hunan Provincial Innovation Foundation for Postgraduate (Grant No. CX20200614).

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/fimmu.2023.1160397/full#supplementary-material

References

1. Schultz KT, Grieder F. Structure and function of the immune system. *Toxicol Pathol* (1987) 15(3):262–4. doi: 10.1177/019262338701500301

2. Tao JH, Cheng M, Tang JP, Liu Q, Pan F, Li XP. Foxp3, regulatory T cell, and autoimmune diseases. *Inflamm* (2017) 40(1):328–39. doi: 10.1007/s10753-016-0470-8

3. Abdolmaleki F, Kovanen PT, Mardani R, Gheibi-Hayat SM, Bo S, Sahebkar A. Resolvins: emerging players in autoimmune and inflammatory diseases. *Clin Rev Allergy Immunol* (2020) 58(1):82–91. doi: 10.1007/s12016-019-08754-9

4. Cooper GS, Bynum ML, Somers EC. Recent insights in the epidemiology of autoimmune diseases: improved prevalence estimates and understanding of clustering of diseases. *J Autoimmun* (2009) 33(3-4):197–207. doi: 10.1016/j.jaut.2009.09.008

5. Zhernakova A, van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nat Rev Genet* (2009) 10(1):43–55. doi: 10.1038/nrg2489

6. Richard-Miceli C, Criswell LA. Emerging patterns of genetic overlap across autoimmune disorders. *Genome Med* (2012) 4(1):6. doi: 10.1186/gm305

7. Gandhi GR, Jothi G, Mohana T, Vasconcelos ABS, Montalvão MM, Hariharan G, et al. Anti-inflammatory natural products as potential therapeutic agents of rheumatoid arthritis: a systematic review. *Phytomedicine* (2021) 93:153766. doi: 10.1016/j.phymed.2021.153766

8. Terao C, Ikari K, Nakayamada S, Takahashi Y, Yamada R, Ohmura K, et al. A twin study of rheumatoid arthritis in the Japanese population. *Modern Rheumatol* (2016) 26(5):685–9. doi: 10.3109/14397595.2015.1135856

9. MacGregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, Aho K, et al. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheumatism* (2000) 43(1):30–7. doi: 10.1002/1529-0131(200001)43:1<30::AID-ANR5>3.0.CO;2-B

10. Hemminki K, Li X, Sundquist J, Sundquist K. Familial associations of rheumatoid arthritis with autoimmune diseases and related conditions. *Arthritis Rheumatism* (2009) 60(3):661–8. doi: 10.1002/art.24328

11. Kuo CF, Grainge MJ, Valdes AM, See LC, Yu KH, Shaw SWS, et al. Familial aggregation of rheumatoid arthritis and co-aggregation of autoimmune diseases in affected families: a nationwide population-based study. *Rheumatology* (2017) 56(6):928–33. doi: 10.1093/rheumatology/kew500

12. Suzuki A, Kochi Y, Okada Y, Yamamoto K. Insight from genome-wide association studies in rheumatoid arthritis and multiple sclerosis. *FEBS Lett* (2011) 585(23):3627–32. doi: 10.1016/j.febslet.2011.05.025

13. Tseng CC, Chang SJ, Tsai WC, Ou TT, Wu CC, Sung WY, et al. Increased incidence of rheumatoid arthritis in multiple sclerosis: a nationwide cohort study. *Medicine* (2016) 95(26):e3999. doi: 10.1097/MD.0000000000003999

14. Su E, Novic J, Han MH. Emergence of rheumatoid arthritis following exposure to natalizumab. *Multiple Sclerosis Related Disord* (2020) 40:101936. doi: 10.1016/j.msard.2020.101936

15. Chen Y, Chen L, Xing C, Deng G, Zeng F, Xie T, et al. The risk of rheumatoid arthritis among patients with inflammatory bowel disease: a systematic review and meta-analysis. *BMC Gastroenterol* (2020) 20(1):192. doi: 10.1186/s12876-020-01339-3

16. Meisinger C, Freuer D. Rheumatoid arthritis and inflammatory bowel disease: a bidirectional two-sample mendelian randomization study. *Semin Arthritis Rheumatism* (2022) 55:151992. doi: 10.1016/j.semarthrit.2022.151992

17. Yang Y, Musco H, Simpson-Yap S, Zhu Z, Wang Y, Lin X, et al. Investigating the shared genetic architecture between multiple sclerosis and inflammatory bowel diseases. *Nat Commun* (2021) 12(1):5641. doi: 10.1038/s41467-021-25768-0

18. Marrie RA, Walld R, Bolton JM, Sareen J, Patten SB, Singer A, et al. Uptake of influenza vaccination among persons with inflammatory bowel disease, multiple sclerosis or rheumatoid arthritis: a population-based matched cohort study. *CMAJ Open* (2021) 9(2):E510–21. doi: 10.9778/cmajo.20200105

19. Marrie RA, Walker JR, Graff LA, Patten SB, Bolton JM, Marriott JJ, et al. Gender differences in information needs and preferences regarding depression among individuals with multiple sclerosis, inflammatory bowel disease and rheumatoid arthritis. *Patient Educ Couns* (2019) 102(9):1722–9. doi: 10.1016/j.pec.2019.04.007

20. Blair M. Diabetes mellitus review. *Urol Nurs* (2016) 36(1):27–36. doi: 10.7257/1053-816X.2016.36.1.27

21. Syed FZ. Type 1 diabetes mellitus. *Ann Internal Med* (2022) 175(3):ITC33–48. doi: 10.7326/AITC202203150

22. Lillegraven S, Greenberg JD, Reed GW, Saunders K, Curtis JR, Harrold L, et al. Immunosuppressive treatment and the risk of diabetes in rheumatoid arthritis. *PloS One* (2019) 14(1):e0210459. doi: 10.1371/journal.pone.0210459

23. Verma AK, Bhatt D, Goyal Y, Dev K, Beg MMA, Alsahli MA, et al. Association of rheumatoid arthritis with diabetic comorbidity: correlating accelerated insulin resistance to inflammatory responses in patients. *J Multidiscip Healthc* (2021) 14:809–20. doi: 10.2147/JMDH.S285469

24. Almeida C, Venade G, Duarte D, Vaz A, Nascimento E. Type 1 diabetes mellitus and multiple sclerosis: an association to consider. *Cureus* (2022) 14(10):e30762. doi: 10.7759/cureus.30762

25. Ahmad R, Ahsan H. Dual autoimmune diseases: rheumatoid arthritis with systemic lupus erythematosus and type 1 diabetes mellitus with multiple sclerosis. *Rheumatol Autoimmun* (2022) 2(3):120–8. doi: 10.1002/rai2.12037

26. Andersen V, Pedersen AK, Möller S, Green A. Chronic inflammatory diseases - diabetes mellitus, rheumatoid arthritis, coeliac disease, crohn's disease, and ulcerative colitis among the offspring of affected parents: a Danish population-based registry study. *Clin Epidemiol* (2021) 13:13–20. doi: 10.2147/CLEP.S286623

27. Zhao C, Jia X, Wang Y, Luo Z, Fan J, Shi X, et al. Overlapping genetic susceptibility of seven autoimmune diseases:SPU tests based on genome-wide association summary statistics. *Gene* (2022) 851:147036. doi: 10.1016/j.gene.2022.147036

28. Coenen MJ, Trynka G, Heskamp S, Franke B, van Diemen CC, Smolonska J, et al. Common and different genetic background for rheumatoid arthritis and coeliac disease. *Hum Mol Genet* (2009) 18(21):4195–203. doi: 10.1093/hmg/ddp365

29. Martin JE, Assassi S, Diaz-Gallo LM, Broen JC, Simeon CP, Castellvi I, et al. A systemic sclerosis and systemic lupus erythematosus pan-meta-GWAS reveals new shared susceptibility loci. *Hum Mol Genet* (2013) 22(19):4021–9. doi: 10.1093/hmg/ddt248

30. López-Isac E, Martín JE, Assassi S, Simeón CP, Carreira P, Ortego-Centeno N, et al. Brief report: IRF4 newly identified as a common susceptibility locus for systemic sclerosis and rheumatoid arthritis in a cross-disease meta-analysis of genome-wide association studies. *Arthritis Rheumatol* (2016) 68(9):2338–44. doi: 10.1002/art.39730

31. Márquez A, Vidal-Bralo L, Rodríguez-Rodríguez L, González-Gay MA, Balsa A, González-Álvaro I, et al. A combined large-scale meta-analysis identifies COG6 as a novel shared risk locus for rheumatoid arthritis and systemic lupus erythematosus. *Ann Rheumatic Dis* (2017) 76(1):286–94. doi: 10.1136/annrheumdis-2016-209436

32. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK biobank. *Nat Genet* (2018) 50(11):1593–9. doi: 10.1038/s41588-018-0248-z

33. Crouch DJM, Inshaw JRJ, Robertson CC, Zhang JY, Chen WM, Onengut-Gumuscu S, et al. Enhanced genetic analysis of type 1 diabetes by selecting variants on both effect size and significance, and by integration with autoimmune thyroid disease. *BioRxiv* (2021). doi: 10.1101/2021.02.05.429962

34. Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS catalog: knowledgebase and deposition resource. *Nucleic Acids Res* (2022) 51:D977–85. doi: 10.1093/nar/gkac1010

35. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang JSchizophrenia Working Group of the Psychiatric Genomics Consortium, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* (2015) 47(3):291–5. doi: 10.1038/ng.3211

36. Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature* (2015) 526(7571):68–74. doi: 10.1038/nature15393

37. Shi H, Mancuso N, Spendlove S, Pasaniuc B. Local genetic correlation gives insights into the shared genetic architecture of complex traits. *Am J Hum Genet* (2017) 101(5):737–51. doi: 10.1016/j.ajhg.2017.09.022

38. Zhu X, Feng T, Tayo BO, Liang J, Young JH, Franceschini N, et al. Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *Am J Hum Genet* (2015) 96(1):21–36. doi: 10.1016/j.ajhg.2014.11.011

39. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* (2007) 81(3):559–75. doi: 10.1086/519795

40. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PloS Comput Biol* (2015) 11(4):e1004219. doi: 10.1371/journal.pcbi.1004219

41. Gerring ZF, Mina-Vargas A, Gamazon ER, Derks EM. E-MAGMA: an eQTL-informed method to identify risk genes using genome-wide association study summary statistics. *Bioinformatics* (2021) 37(16):2245–9. doi: 10.1093/bioinformatics/btab115

42. Consortium. The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science* (2020) 369(6509):1318–30. doi: 10.1126/science.aaz1776

43. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* (2016) 44(W1):W90–7. doi: 10.1093/nar/gkw377

44. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* (2015) 43:D447–52. doi: 10.1093/nar/gku1003

45. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nat Genet* (2018) 50(5):693–8. doi: 10.1038/s41588-018-0099-7

46. An J, Gharahkhani P, Law MH, Ong JS, Han X, Olsen CM, et al. Gastroesophageal reflux GWAS identifies risk loci that also associate with subsequent severe esophageal diseases. *Nat Commun* (2019) 10(1):4219. doi: 10.1038/s41467-019-11968-2

47. Peyrot WJ, Price AL. Identifying loci with different allele frequencies among cases of eight psychiatric disorders using CC-GWAS. *Nat Genet* (2021) 53(4):445–54. doi: 10.1038/s41588-021-00787-1

48. Saevarsdottir S, Stefansdottir L, Sulem P, Thorleifsson G, Ferkingstad E, Rutsdottir G, et al. Multiomics analysis of rheumatoid arthritis yields sequence variants that have large effects on risk of the seropositive subset. *Ann Rheumatic Dis* (2022) 81(8):1085–95. doi: 10.1136/annrheumdis-2021-221754

49. Andlauer TF, Buck D, Antony G, Bayas A, Bechmann L, Berthele A, et al. Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation. *Sci Adv* (2016) 2(6):e1501678. doi: 10.1126/sciadv.1501678

50. Jiang X, Källberg H, Chen Z, Ärlestig L, Rantapää-Dahlqvist S, Davila S, et al. An immunochip-based interaction study of contrasting interaction effects with smoking in ACPA-positive versus ACPA-negative rheumatoid arthritis. *Rheumatology* (2016) 55 (1):149–55. doi: 10.1093/rheumatology/kev285

51. Kugathasan S, Baldassano RN, Bradfield JP, Sleiman PM, Imielinski M, Guthery SL, et al. Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease. *Nat Genet* (2008) 40(10):1211–5. doi: 10.1038/ng.203

52. Tomer Y, Dolan LM, Kahaly G, Divers J, D'Agostino RBJr., Imperatore G, et al. Genome wide identification of new genes and pathways in patients with both autoimmune thyroiditis and type 1 diabetes. *J Autoimmun* (2015) 60:32–9. doi: 10.1016/j.jaut.2015.03.006

53. Syreeni A, Sandholm N, Sidore C, Cucca F, Haukka J, Harjutsalo V, et al. Genome-wide search for genes affecting the age at diagnosis of type 1 diabetes. *J Internal Med* (2021) 289(5):662–74. doi: 10.1111/joim.13187

54. Baranova A, Cao H, Zhang F. Causal associations and shared genetics between hypertension and COVID-19. *J Med Virol* (2023) 95(4):e28698. doi: 10.1002/jmv.28698

55. Hakonarson H, Grant SF, Bradfield JP, Marchand L, Kim CE, Glessner JT, et al. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature* (2007) 448:591–4. doi: 10.1038/nature06010

56. Onengut-Gumuscu S, Chen WM, Burren O, Cooper NJ, Quinlan AR, Mychaleckyj JC, et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat Genet* (2015) 47(4):381–6. doi: 10.1038/ng.3245

57. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet* (2009) 41(6):703–7. doi: 10.1038/ng.381

58. Robertson CC, Inshaw JRJ, Onengut-Gumuscu S, Chen WM, Santa Cruz DF, Yang H, et al. Fine-mapping, trans-ancestral and genomic analyses identify causal variants, cells, genes and drug targets for type 1 diabetes. *Nat Genet* (2021) 53(7):962–71. doi: 10.1038/s41588-021-00880-5

59. Laufer VA, Tiwari HK, Reynolds RJ, Danila MI, Wang J, Edberg JC, et al. Genetic influences on susceptibility to rheumatoid arthritis in African-americans. *Hum Mol Genet* (2019) 28:858–74. doi: 10.1093/hmg/ddy395

60. Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet* (2021) 53(10):1415–24. doi: 10.1038/s41588-021-00931-x

61. Liao KP, Gunnarsson M, Källberg H, Ding B, Plenge RM, Padyukov L, et al. Specific association of type 1 diabetes mellitus with anti-cyclic citrullinated peptide-positive rheumatoid arthritis. *Arthritis Rheumatol* (2009) 60(3):653–60. doi: 10.1002/art.24362

62. Hellwege JN, Velez Edwards DR, Giri A, Qiu C, Park J, Torstenson ES, et al. Mapping eGFR loci to the renal transcriptome and phenome in the VA million veteran program. *Nat Commun* (2019) 10(1):3842. doi: 10.1038/s41467-019-11704-w

63. Mero IL, Gustavsen MW, Sæther HS, Flåm ST, Berg-Hansen P, Søndergaard HB, et al. Oligoclonal band status in Scandinavian multiple sclerosis patients is associated with specific genetic risk alleles. *PloS One* (2013) 8(3):e58352. doi: 10.1371/journal.pone.0058352

64. Orozco G, Viatte S, Bowes J, Martin P, Wilson AG, Morgan AW, et al. Novel rheumatoid arthritis susceptibility locus at 22q12 identified in an extended UK genome-wide association study. *Arthritis Rheumatol* (2014) 66(1):24–30. doi: 10.1002/art.38196

65. Yadav P, Ellinghaus D, Rémy G, Freitag-Wolf S, Cesaro A, Degenhardt F, et al. Genetic factors interact with tobacco smoke to modify risk for inflammatory bowel disease in humans and mice. *Gastroenterology* (2017) 153(2):550–65. doi: 10.1053/j.gastro.2017.05.010

66. Zhao W, Rasheed A, Tikkanen E, Lee JJ, Butterworth AS, Howson JMM, et al. Identification of new susceptibility loci for type 2 diabetes and shared etiological pathways with coronary heart disease. *Nat Genet* (2017) 49(10):1450–7. doi: 10.1038/ng.3943

67. Asano K, Matsushita T, Umeno J, Hosono N, Takahashi A, Kawaguchi T, et al. A genome-wide association study identifies three new susceptibility loci for ulcerative colitis in the Japanese population. *Nat Genet* (2009) 41(12):1325–9. doi: 10.1038/ng.482

68. Suzuki K, Akiyama M, Ishigaki K, Kanai M, Hosoe J, Shojima N, et al. Identification of 28 new susceptibility loci for type 2 diabetes in the Japanese population. *Nat Genet* (2019) 51(3):379–86. doi: 10.1038/s41588-018-0332-4

69. Andlauer TFM, Link J, Martin D, Ryner M, Hermanrud C, Grummel V, et al. Treatment- and population-specific genetic risk factors for anti-drug antibodies against interferon-beta: a GWAS. *BMC Med* (2020) 18(1):298. doi: 10.1186/s12916-020-01769-6

70. Zhao J, Cai B, Shao Z, Zhang L, Zheng Y, Ma C, et al. TRIM26 positively regulates the inflammatory immune response through K11-linked ubiquitination of TAB1. *Cell Death Differentiation* (2021) 28(11):3077–91. doi: 10.1038/s41418-021-00803-1

71. Zhao YY, Yan DJ, Chen ZW. Role of AIF-1 in the regulation of inflammatory activation and diverse disease processes. *Cell Immunol* (2013) 284(1-2):75–83. doi: 10.1016/j.cellimm.2013.07.008

72. Salao K, Jiang L, Li H, Tsai VW, Husaini Y, Curmi PM, et al. CLIC1 regulates dendritic cell antigen processing and presentation by modulating phagosome acidification and proteolysis. *Biol Open* (2016) 5(5):620–30. doi: 10.1242/bio.018119

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 11 January 2024
DOI 10.3389/fgene.2023.1272912

OPEN ACCESS

EDITED BY
Che Kang Lim,
Singapore General Hospital, Singapore

REVIEWED BY
Intan Juliana Abd Hamid,
University of Science Malaysia (USM), Malaysia
Atıl Bisgin,
Cukurova University AGENTEM (Adana Genetic Diseases Diagnosis and Treatment Center) & InfoGenom RD Laboratories, Türkiye

*CORRESPONDENCE
Charlotte Cunningham-Rundles,
charlotte.cunningham-rundles@mssm.edu

RECEIVED 04 August 2023
ACCEPTED 09 November 2023
PUBLISHED 11 January 2024

CITATION
Cunningham-Rundles C, Casanova J-L and Boisson B (2024), Genetics and clinical phenotypes in common variable immunodeficiency.
Front. Genet. 14:1272912.
doi: 10.3389/fgene.2023.1272912

Genetics and clinical phenotypes in common variable immunodeficiency

Charlotte Cunningham-Rundles[1,2,3]*, Jean-Laurent Casanova[4,5,6,7,8] and Bertrand Boisson[4,5,6]

[1]Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States, [2]Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY, United States, [3]Division of Clinical Immunology, Departments of Medicine and Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY, United States, [4]St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY, United States, [5]Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM, Necker Hospital for Sick Children, Paris, France, [6]Paris Cité Université, Imagine Institute, Paris, France, [7]Department of Pediatrics, Necker Hospital for Sick Children, Paris, France, [8]Howard Hughes Medical Institute, New York, NY, United States

Common variable immunodeficiency (CVID) is one of the most common symptomatic groups of inborn errors of immunity. In addition to infections resulting from insufficient levels of immune globulins and antibodies, many patients develop inflammatory or autoimmune conditions, which are associated with increased mortality. This aspect of CVID has been the focus of many studies, and dissecting the clinical phenotypes of CVID, has had the goal of providing biomarkers to identify these subjects, potentially at the time of diagnosis. With the application of whole exome (WES) and whole genome analyses, an increasing number of monogenic causes of CVID have been elucidated. From the standpoint of the practicing physician, an important question is whether the clinical phenotype, particularly the occurrence of autoinflammation of autoimmunity, might suggest the likelihood of identifying a causative mutation, and if possible the gene most likely to underlie CVID. We addressed this question in a patient group of 405 subjects diagnosed with CVID from one medical center.

KEYWORDS

common variable immunodeficiency, genetics, autoimmunity, lung disease, granulomatous disease, cancer, lymphoma

Introduction

Common variable immunodefciency (CVID) is one of the more frequently encountered immune defects in clinical practice, with an estimated incidence of about 1 in 20,000. The diagnosis is made in a male or female patient with reduced serum levels of IgG, along with IgA, and/or IgM with documented defects of antibody production to both protein and carbohydrate antigens, and the exclusion of other causes of hypogammaglobulinemia such as physiologic immaturity, medications, malignancy, or protein losses (Bonilla et al., 2016; Registry, 2016; Odnoletkova et al., 2018; Seidel et al., 2019). While considered genetic defects, most newly diagnosed patients are between the ages of 20 and 40 years old. (Resnick et al., 2012a; Gathmann et al., 2014; Odnoletkova et al., 2018). One of the clinical features of the CVID syndrome that has emerged is that about half of these patients have infections as the central manifestation, which can be successfully treated or prevented with antibiotics and immunoglobulins. However, the others also have various apparently non-infectious, autoimmune, autoinflammatory, neoplastic and/or lymphoproliferative manifestations,

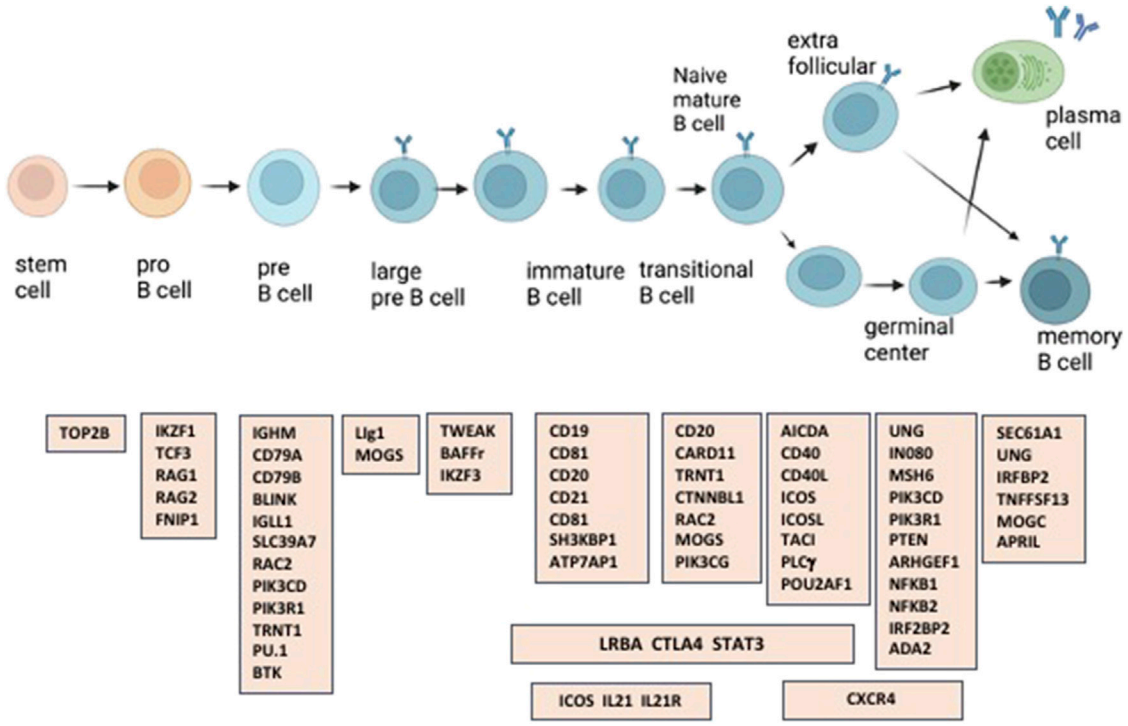


FIGURE 1
A number of the immune defects found in patients with CVID, are in genes involved in the generation and maturation of human B cells.

often associated with systemic immune activation (Wehr et al., 2008; Resnick et al., 2012b; Cols et al., 2016; Smith and Cunningham-Rundles, 2021; Ho and Cunningham-Rundles, 2022). Patients in the second group often have autoimmune or inflammatory features as the initial presentation and primary clinical manifestation, with less obvious susceptibility to significant infectious diseases; these subjects also have increased morbidity and mortality as compared to those with the infection-only phenotype (Chapel et al., 2008; Resnick et al., 2012a). A number of studies have probed reasons for the striking heterogeneity of this CVID patient pool (Wehr et al., 2008; Chapel and Cunningham-Rundles, 2009; Resnick et al., 2012a). These studies have sought biomarkers to identify these subjects, preferably at the time of diagnosis (Ho and Cunningham-Rundles, 2022). Some of these markers include identification of subjects with loss of peripheral isotype switched memory B cells, increased $CD21^{low}$ B cells (<10%), and/or reduced numbers of T cells, especially naïve CD4 T cells (Warnatz et al., 2002; Fevang et al., 2007; Sanchez-Ramon et al., 2008; Wehr et al., 2008; Malphettes et al., 2009; Mouillot et al., 2010). Other markers more recently defined include elevated markers of systemic immune activation: serum lipopolysaccharide binding protein (LBP), sCD14 (Barbosa et al., 2012; Litzman et al., 2012; Fraz et al., 2022) and more recently, serum zonulin and circulating bacterial DNA (Ho et al., 2021). These correlative biomarkers attempt to distinguish many patients with the infection-only clinical phenotype from those with more inflammatory complications, but they do not address the molecular mechanisms.

With the availability of whole exome (WES) and whole genome analyses (WGS), an increasing number of monogenic causes of the CVID phenotype have been elucidated, now accounting for about 25%–30% of subjects (Maffucci et al., 2016; Tuijnenburg et al., 2018; Abolhassani et al., 2020; Ramirez et al., 2021; Rojas-Restrepo et al., 2021). Several recent reports have described the results of genetic analyses of large CVID patient populations, with differences noted due to the location of the populations studied and ethnic background of the patients (Abolhassani et al., 2020; Rojas-Restrepo et al., 2021). The many genes identified in cohorts of subjects diagnosed with CVID, reflect the complex requirements of class switch recombination, B cell antigen signaling, activation, migration, long-term survival, and maturation and retention of antibody-secreting memory B cells into the plasma cell stage (Figure 1). From the standpoint of the practicing physician, an important question is whether the clinical phenotype suggests the possibility of identifying a causative mutation and if so, the gene(s) most likely to underlie the immune defect. Here we address this question in a large patient group from one medical center, encompassing an urban patient population on the East Coast of the United States.

Methods

Patient selection

Subjects were seen in the Immune Deficiency program at the Icahn School of Medicine at Mount Sinai. Patients were diagnosed with CVID using established criteria, including serum IgG and IgA and/or IgM deficiency with proven loss of antibody production

(Bonilla et al., 2016; Picard et al., 2018; Seidel et al., 2019). Immunologic and clinical histories were collected from the clinical record and selected manifestations of inflammatory/autoimmune complications were recorded. For purposes of the current study, these sometimes partly overlapping medical conditions are autoimmunity, interstitial lung disease, granulomatous disease identified in one or more tissues, cancer, lymphoma, significant gastrointestinal disease, and previous splenectomy. Ethical permission for these studies was obtained from the Mount Sinai Institutional Review Board, and Informed consent was obtained from all individuals and/or their legal guardians.

TABLE 1 CVID subjects.

Parameters	N
Number	405
Males	218
Females	187
Median age, years (range)	44 (5–77)
Relatives with immune defects	26
Gene candidates identified (%)	128 (31%)

Genetic evaluation

Whole exome sequencing: Genetic evaluation was done by whole exome sequencing (WES) as previously described (Maffucci et al., 2016; Picard et al., 2018; Maffucci et al., 2019). For this genomic DNA was extracted from peripheral blood mononuclear cells and sheared with a Covaris S2 Ultrasonicator. An adaptor-ligated library was prepared with the Paired-End Sample Prep kit V1 (Illumina). Exome capture was performed with the SureSelect Human All Exon kit (Agilent Technologies). Massively parallel sequencing was performed on a HiSeq 2,500 (Illumina), which generates 100-base reads. Sequences were aligned for variant calling and annotation with the human genome reference sequence (hg19 build) using BWA aligner (Li and Durbin, 2009). Downstream processing was performed with the Genome analysis toolkit (GATK) (McKenna et al., 2010), SAMtools (Li et al., 2009), and Picard Tools (http://picard.sourceforge.net/). A GATK UnifiedGenotyper and a GATK IndelGenotyperV2 were used to identify substitution and indel variant calls, respectively. Calls with a read coverage of ≤2x and a Phred-scaled single-nucleotide polymorphism (SNP) quality of ≤20 were filtered out. All variants were annotated with the GATK Genomic Annotator (Broad Institute). Heterozygous and homozygous variants were excluded if the allele frequencies in the general population were greater than 0.01% or 1.0%, respectively, in the Exome Aggregation Consortium database (ExAC, Broad Institute) and Genome Aggregation Database (gnomAD, Broad Institute) in respect with genetic model tested. This filtering strategy excluded polymorphic variants from consideration. Familial segregation was studied when samples were available. Other candidate mutations were confirmed by examining read alignment in the Integrated Genomics Viewer (IGV; Broad Institute). All confirmed mutations were subsequently analyzed using computational predictors of mutation severity including Sift (Ng and Henikoff, 2003), Polyphen 2 (Adzhubei et al., 2010). and Combined Annotation Dependent Depletion (CADD) (Kircher et al., 2014) and were compared with the gene-specific mutation significance cutoff (MSC) (Itan et al., 2016). Variants with CADD scores below the gene-specific MSC were excluded with the exception of *CXCR4*, *LIG1*, *LRBA* and *NFKB2*, 4 genes with a very high MSC (>32), but known to be causal of CVID. Confirmed variations were also screened through the Human Gene Mutation Database (Stenson et al., 2003) to identify published disease-associated variations. In a number of cases, the variants selected were tested for pathogenicity, and if not, categorized (as likely benign, VUS - Variant of Uncertain Significance, or likely pathogenic.). The pathogenicity of all disease attributable gene variants was evaluated using the updated guideline for interpretation of molecular sequencing by the American College of Medical Genetics and Genomics (ACMG) considering the allele frequency, computational data, immunological/functional data, familial segregation and parental data and clinical phenotyping (Richards et al., 2015).

Using a targeting panel of genes: More recently, patient exomes were also examined for mutations in 429 genes associated with a primary immune deficiency disease (Invitae Diagnostics), when faster results were needed and insurance or other payment for this service was available. For inheritance questions, the targeted sequence method was used in particular as it has the Clinical Laboratory Improvement Amendments (CLIA) certification needed for clinical decision analysis. As the targeted panel contains only a defined panel of genes, the above WES method was also used on these same samples so that the data could be verified, and so that additional queries for new genes could be undertaken at a later date.

Statistics

For evaluating the significance of genetics as related to clinical observations, Graphpad Prism was used; a *p*-value of less than 0.05 was considered significant.

Results

Patients

The Mount Sinai cohort of 405 genetically-tested CVID patients included 26 cases in which a family history was noted (Table 1). The age range of patients was 5–77 years, with median age of 44 years; 187 were female and 218 were male. Of these subjects, most were of European descent, 22 were Hispanic, 12 were Black, and 4 were Asian.

Mutations identified

In this cohort, 125 of 405 subjects (31%) had mutations considered deleterious while the remainder (280) had no clear genetic abnormality identified. As reported previously, around 10% of our patients (42 subjects) have variants in the *TNFRSF13B/TACI* gene (Transmembrane Activator and CAML Interactor) most of these known to be functionally deleterious

TABLE 2 Gene variants identified - 125 subjects (31% of the group).

Gene variants	Number	Name	Inheritance
TNFRSF13B[a]	42	Transmembrane Activator and CAML Interactor	AD
NFKB1	16	Nuclear Factor Kappa B Subunit 1	AD
NFKB2	7	Nuclear Factor Kappa B Subunit 2	AD
IRF2BP2	6	Interferon regulatory factor-2 binding **protein**	AD
CTLA4	6	Cytotoxic T-Lymphocyte Associated Protein 4	AD
IKZF1	6	IKAROS Family Zinc Finger 1	AD
TCF3	5	Transcription Factor 3	AD
BACH2	5	BTB Domain and CNC Homolog 2	AD
KMT2D	4	Lysine Methyltransferase 2	AD
STAT3	3	Signal Transducer And Activator Of Transcription 3	AD
PIK3CD	3	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Delta	AD
LRBA	3, compd het	LPS Responsive Beige-Like Anchor Protein	AR
CXCR4	3	C-X-C Motif Chemokine Receptor 4	AD
DiGeorge 22q11 or *TBX1*	3	DiGeorge syndrome	AD
WAS	2	Wiskott Aldrich syndrome	XL
RAG1/RAG2	2	Recombination Activating Genes 1/2	AR
AICDA	1, homozygous	Activation induced cytidine deaminase	AR
STXBP2	1, compd het	Syntaxin Binding Protein 2	AR
PMM2	1, compd het	Phosphomannomutase	AR
PIK3R1	1	Phosphoinositide-3-Kinase Regulatory Subunit 1	AD/AR
LIG4	1 homozygous	DNA Ligase 4	AR
JAK1	1	Janus Kinase 1	AR
IKBKG	1	Regulatory gamma subunit of the IKB kinase (IKK)	XL
TBX1	1	T-box protein 1	AD
PMS2	1, compd het	PMS1 Homolog 2, Mismatch Repair System	AR
FOXP3	1	Fork-winged helix family	XL
LIG1	1, compd het	DNA Ligase 1	AR
CIITA	1, compd het	Master regulator of MCH class II gene transcription	AR
BTK	1	Bruton Tyrosine Kinase	XL
ADA2	1	Adenosine deaminase 2	AR
CD40L	1	CD40 Ligand	XL
RTEL1	1, compd het	regulator of telomere elongation helicase 1	AR

[a]Includes compound heterozygous mutations in 4; homozygous mutations in one; 8 of these also had an additional deleterious mutation in *TBX1, TCF3, IL10RA, NFKB2, NBAS,* or a DiGeorge chromosome loss; AD , autosomal dominant; AR , autosomal recessive; XL = X linked.

(Salzer and Grimbacher, 2021). These included compound heterozygous mutations in 5 subjects, and homozygous mutations in one subject. These variants would be considered associated but not causative of CVID. (Table 2). Sixteen other subjects had autosomal dominant (AD) *NFKB1* deficiency (Nuclear Factor Kappa B Subunit 1), and 7 others had *NFKB2* (Nuclear Factor Kappa B Subunit 2) defects, both viewed as causative of this immune defect (Chen et al., 2013; Tuijnenburg et al., 2018; Li et al., 2021). More than one subject had mutations in heterozygous genes previously found in subjects with a CVID phenotype: *IRF2BP2, CTLA4 and IKZF1* (in 6 subjects each), *TCF3* (in 5 subjects), *BACH2* (in 4 subjects), and in *STAT3, and PIK3CD* (3 subjects for each). Three other subjects had autosomal recessive (AR) deficiency of *LRBA*. Four adult subjects with infections, autoimmunity and mild

TABLE 3 Subjects with more than one gene variant.

Subject	Variant		Other variants	
1	*BACH2*	p.Gly483Ser	*POLE*	c.2706 + 1G>T
2	*BACH2*	p.Glu797*	*IKZF1*	p.Asn350His
3	*CTLA4*	c.109 + 1G>A	*IRFBP2*	p.Gln97His
4	*JAK1*	p.Asn76Ser	*STAT3*	p.Val461Leu
5	*NFKB2*	p.Gly719Glu	*TACI*	p.Leu69Thr fs*12
6	*NFKB2*	splicing	*TACI*	p.Cys104Arg
7	*PIK3R1*	start_gained/start_gained	*DCLRE1C*	del exon 1–3
8	*TACI*	p. Ala181Glu	*TBX1*	p.Leu1007Profs*2
9	*TACI*	p.Ala181Glu	Di George	22q11.2 deletion
10	*TACI*	p.Glu236*	*PMM2*	p.Arg141His
11	*TCF3*	p.Asn554Ser	*DOCK8*	c.54-1G>T
12	*TCF3*	p.Ile562Val	*TACI*	p.Leu69fs/p. Cys104Arg
13	*TCF3*	p.Pro96Leu	*TACI*	p.Lys188del
14	*TACI*	p.Ala181Glu	*IL10RA*	p.Arg147Profs*4
			NBAS	Deletion (Exons 45–52) p.Ser712*)
			TMPRSS15	
15	*RAG1*	p.Asp212Asn	*RAG2*	p.Asp400His
16	*RAG1*	p.Asn968Lys	*RAG2*	p.Met110Leu

retardation had mutations in *KMT2D* (Lysine Methyltransferase 2), a gene associated with Kabuki syndrome. Two sisters and the son of one of them, with no warts and moderate neutropenia, but severe autoimmune thrombocytopenia and autoimmune hemolytic anemia, had frameshift mutations in *CXCR4* (C-X-C Motif Chemokine Receptor 4), a gene associated with WHIM syndrome (warts, hypogammaglobulinemia, infections, and myelokathexis) (Maffucci et al., 2016; Abolhassani et al., 2020). Note that as in other reports, genes identified with previously un-identified X-immune linked defects were also noted in this cohort, *BTK, CD40L, IKBKG* and *WAS* (Table 2). Complicating the genetics is that in 15 subjects studied, more than one heterozygous variant, predicted to be deleterious, was identified. This included 8 subjects with at least one *TACI* variants, but additional variants in other autosomal dominant or recessive genes were also noted, for example, in *TBX1, TCF3, IL10-RA, NFKB2, NBAS, RAG1, RAG2* or a DiGeorge chromosome 22q deletion (Table 3). Further information related to minor allele frequency, and predicted deleterious effects on the selected variants are included in Methods and Supplementary Table S1.

Genetics and clinical phenotypes

We then examined if subjects with autoinflammatory, autoimmune, lymphoproliferative, neoplastic, granulomatous infiltrates, and/or gastrointestinal complications were more likely to have mutations in one or more of the genes identified in CVID, in contrast to others for whom a gene was not identified. Table 4 outlines the results for this cohort, dividing subjects according to whether or not a gene defect was identified in subjects with autoimmunity, significant pulmonary, gastrointestinal disease or liver disease, biopsy-demonstrated granulomatous disease, previous splenectomy, cancer, or lymphoma. Various forms of autoimmunity were noted in 151 subjects (37%) of the 405 group genetically tested, with no sex predominance. Of the 125 subjects with genes identified, 59 (47%) had autoimmunity, while of the larger group of 281 subjects with no gene noted, 32% had autoimmunity, suggesting some enrichment of this clinical feature in those with any known gene association; however these differences were not statistically significant. We also examined the types of autoimmunity in each group, in those with or without identified gene variants. Tables 5, 6 show these data. However, for both sets of patients, the most prominent autoimmune conditions were cytopenias, particularly thrombocytopenia, hemolytic anemia (or both, *i.e.*, Evan's syndrome) or, in fewer numbers, neutropenia (Figure 2). The mutations found in those with autoimmunity are included in Supplementary Table S2. Note that of these, 14 had mutations in the *TACI* gene (2 with compound heterozygous variants, one with homozygosity) and 4 others were in association with additional deleterious variants). Five of the subjects with autoimmunity had mutations in *NFKB1*, 4 had variants in *CTLA4, IKZF1,* or in *STAT3,* and with other genes noted in other subjects.

Significant pulmonary disease (interstitial lung disease, numerous nodules, and/or known granulomatous disease or

TABLE 4 Percentage of complications in each group.

	Auto-immunity %	Pulmonary %	Gastro-intestinal %	Liver %	Granuloma %	Splenec-tomy %	Cancer %	Lymphoma %
With gene N= 125	47	31	16	17	17.6	13.6	8	6.5
no gene N= 280	32	17.5	18	9	7	6	4.3	7.5

TABLE 5 Autoimmunity: Gene identified.

N = 125		
Condition	Number	%
ITP	25	37
AIHA ITP	16	24
AIHA	4	6
Neutropenia	4	6
Diabetes Mellitus	4	6
Autoimmune hepatitis	3	4
Alopecia	2	3
Pancytopenia	2	3
Uveitis	2	3
Opsoclonus myoclonus	1	1
Psoriatic arthritis	1	1
Myasthenia Gravis	1	1
TTP	1	1
PSC	1	1

respiratory failure) was observed in 39 subjects. Of these (31%) had an identified gene defect, as compared to 17.5% of those with no gene noted, suggesting a slight but not significant enrichment of significant respiratory disease in those with an identified gene, p = 0.053. Of those with significant lung disease, 12 subjects had a *TACI* variant, 4 had *NFKB2,* 4 had *NFKB1* and 3 others had *PI3KCD* variants. Mutations in the genes, BACH2, *KMT2D, LRBA, LIG4, STAT3 FOXP3* and *CASP8* were identified in other patients as outlined (Supplementary Table S3). Thirteen of these subjects had been diagnosed as having granulomatous lymphocytic infiltrates in the lungs (GLILD) (Bates et al., 2004). Overall, the tissue diagnosis of granulomatous disease (in lymph nodes, lung skin, or liver) was noted in 22 (17.7%) of the subjects with defects in known genes (B*ACH2, CTLA4, KMT2D, LRBA, PIK3R1, STAT3, TACI, TCF3 or WAS.*) This was significantly different from the 20 other subjects (7.5%) with granulomatous infiltrations) with no known gene defect, (p = 0.046).

The diagnosis of cancer or lymphoma was made in 18 patients (14.5%) with an identified gene defect, including 9 lymphoid malignancies and 6 other cancers, with the genetic changes noted for each (Table 7). Cancer or lymphoma was also diagnosed in 32 other subjects (11%) with no known genetic variants (not significantly different.) Gastrointestinal and/or liver disease were noted in similar proportions in each group, with the genes noted in Supplementary Table S4. Splenectomy, usually done for treatment of cytopenias, had been performed in 34 patients overall, with 17 (13.7%) of these in subjects with known genetic variants, as opposed to 17 others (6%) with no known genetic association (not significantly different.) Of the splenectomized subjects, 7 had TACI gene variants; others included *NFKB1* in 4, *LRBA, TCF3, STAT3, NFKB2, CTLA4* and *RAG1/RAG2.*

We also considered if those with genetic variants might have specific types of unusual infections. However, a history of infections with *Herpes Zoster*, *Candida* sp, Giardia, *Clostridia difficile*, *Helicobacter pylori*, Norovirus, *Campylobacter*, *Herpes simplex*, or more unusual infections with atypical mycobacteria, *mycoplasma*, histoplasmosis, or cryptococcus, were seen in subjects with and without genetic variants.

Discussion

A number of previous studies have outlined the clinical phenotypes of large patient groups with antibody deficiency who have mutations in selected genes, including the *TACI* gene (Salzer et al., 2005; Zhang et al., 2007; Salzer et al., 2009), *CTLA4* (Schwab et al., 2018), *NFKB1* (Lorenzini et al., 2020), *NFKB2* (Klemann et al., 2019), *STAT3* (Fabre et al., 2019), *PI3KCD* (Jamee et al., 2020), or *LRBA* (Habibi et al., 2019). These studies describe the infectious, autoimmune and inflammatory characteristics of patients with these specific inborn errors of immunity. Here, the genetic analyses of CVID patients from one large cohort were examined to determine if the clinical complications might lead the physician to suspect a genetic defect in one or more of the previously established causal genes. In the current cohort of 405 subjects, 125 (31%) had an identifiable causative or associated genetic variant, similar to other investigated cohorts (Abolhassani et al., 2020; Rojas-Restrepo et al., 2021), however leaving the majority of patients without a known genetic cause. Autoimmunity was one of the commonest conditions for this group of patients with 37% of the patients having one or more of these conditions. Of these, 47% of them carried a predicted deleterious variant, while for those with no gene noted, 32% had autoimmunity. The autoimmune conditions noted were similar for each group, with cytopenias being the most prevalent manifestation, resulting in splenectomy in a number of subjects. While patients with mutations in *CTLA4, IKZF1, STAT3* and *LRBA* were in the autoimmune group, the largest number had variants in the *TACI* gene. While variants in

TABLE 6 Autoimmunity: No gene identified.

N = 280		
Condition	Number	%
ITP	42	45
AIHA ITP	11	12
Diabetes	7	8
AIHA	5	5
Thyroiditis	4	5
Neutropenia	3	3
Uveitis	3	3
RA	3	3
Psoriasis	2	2
Vitiligo	2	2
Multiple Sclerosis	2	2
ANA+	1	1
B12 Deficient	1	1
Anti-phospho-lipid antibody	1	1
Scleroderma	1	1
Red cell aplasia	1	1

AIHA , autoimmune hemolytic anemia; ITP , immune thrombocytopenia; TTP, thrombotic thrombocytopenic purpura; PSC , primary sclerosing cholangitis; ANA, antinuclear antibody; RA , rheumatoid arthritis; PSC , primary sclerosing cholangitis.

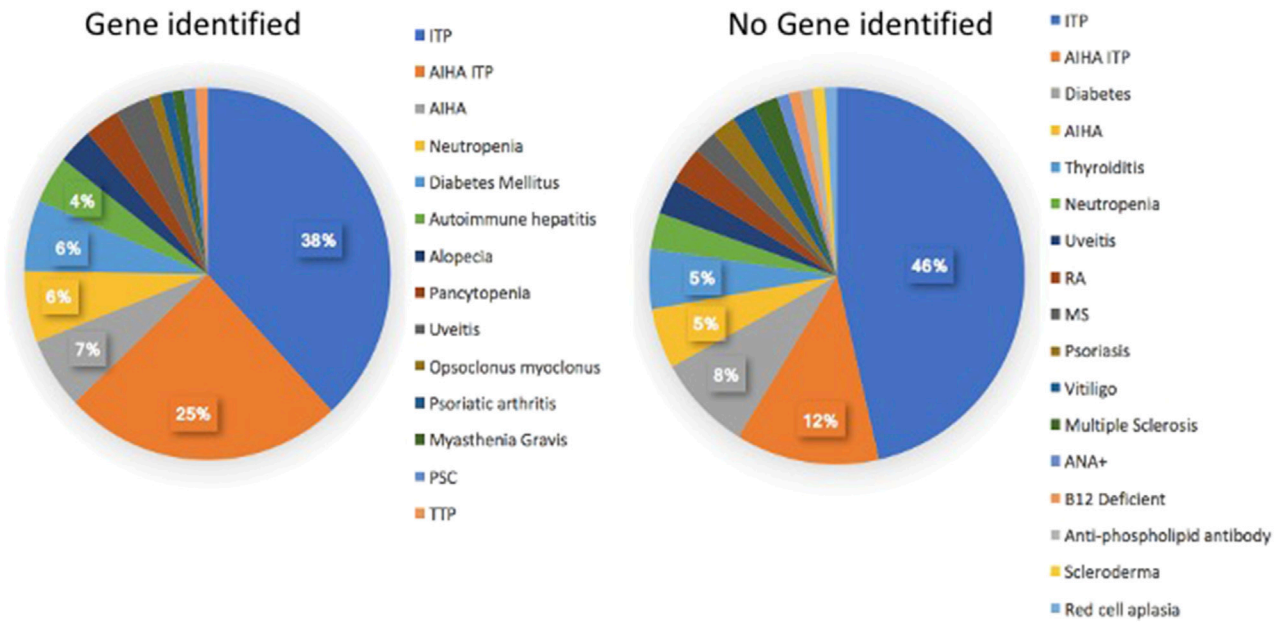


FIGURE 2
The kinds of autoimmune conditions found in patients with, and without identified genetic defects are similar.

TABLE 7 Gene defects in cancers and lymphoma.

Gene	Consequence	Cancer	Lymphoma/Leukemia
BTK	p.Tyr418His	Esophagus	-
CD40L	indel-frameshift	Bladder	-
IKZF1	p.Ser385*		T cell leukemia
LRBA	p.Ile2232Thr/p.Ala892Thr	Mouth	-
NFKB2	p.His98Asn	-	Gastric Maltoma
TACI	p.Cys104Arg		
NFKB2	p.Gly719Glu	-	Gastric Maltoma
TACI	p.Leu69fs		
PI3KCD	p.Glu1021Lys	-	Lymphoma
PI3KCD	p.Glu1021Lys	Ovary	-
PIK3R1	start gained/start gained		MALT Lymphoma
DCLRE1C	del exon 1–3		
RAB27A.	del exon 2	Gall Bladder	
PMS2	p.Ile18Val/p.Arg563Leu		
TACI	p.Cys104Arg		Plasmablastic Lymphoma
TACI	p.Cys104Arg	Rectal	-
TACI	p.Ala181Glu	-	MALT Lymphoma
TCF3	p.Ile562Val	-	Lymphoma
TACI	p.Leu69fs/Cys104Arg		
TACI/	p.Ala181Glu		Lymphoma
TMPRSS15	p.Ser712*		
NBAS	deletion exons 45–52		
IL10RA	p.Arg147Pro fs*4		

the *TACI* gene are not thought to be disease causing, they are commonly associated with autoimmunity in CVID (Salzer et al., 2005; Zhang et al., 2007), possibly explaining this enrichment. If subjects with a TACI variant are excluded, 35% of subjects with a known gene were noted to have autoimmunity, more similar to those for whom no gene was identified (at 32%) as illustrated (Figure 3) The autoimmune cytopenia (ITP or AIHA) resulted in splenectomy in 34 patients, and 7 of the 17 with a known gene, had a *TACI* variant. Significant respiratory disease was also more common in those subjects with a known gene defect (at 31%). Excluding the 8 subjects with a *TACI* variant, 25% of these subjects had this complication, more similar to those with no known genetic background (17.5%). Granulomatous disease was overall, significantly enriched in those with genes identified. In this group, 17.6% had this complication if they had *TACI* variants; if these are excluded, the percentage was 14%, as opposed to 7% of those with no noted genes. Other complications, such as gastrointestinal, substantial liver disease or cancer appeared in subjects with and without identified genetic defects in similar numbers (Figure 3).

As the data presented here was gathered over a decade, a question that might arise if the DNA samples tested over time, using WES or the targeted panel, would lead to different results. We did not find this difficulty, but using the two methods led to confirmation. In addition, the accumulation of genetic data obtained by WES on all samples, also allowed for continued surveying for newly reported mutational differences. This allowed for updating as new genes contributing to the CVID phenotype were identified, and allowed all samples to be examined by the same parameters. The targeting panel was particularly useful for rapid analysis or inheritance questions, but even in these cases, WES was still performed. Limitations of this study include the fairly narrow demographic background of the subjects referred, a known reason for genetic differences in any CVID patient population (Abolhassani et al., 2020; Rojas-Restrepo et al., 2021), and the possibility of incomplete clinical information as these data were collected over varying periods of time. A further limitation is that confirmation of the pathogenicity of the identified variants depends on previous publications, *in vitro* assays where available, and the genetic

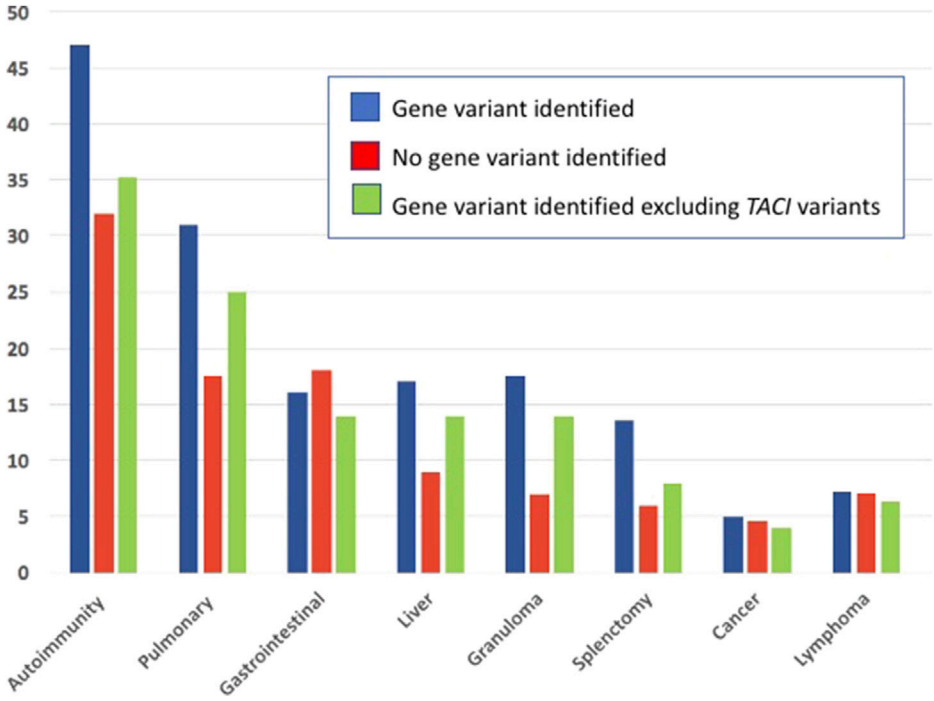


FIGURE 3
The medical complications noted in subjects with identified genetic defects, those with no known gene defects and subjects with a gene variant, but minus those with defects in the TNFRSF13B TACI receptor.

methods used to assign the likelihood of a variant of dexterous variant exerting a deleterious change.

Overall, we conclude that CVID subjects with currently identifiable gene variants, either associated with or causative of this immune defect, appear to have an increased numbers of autoimmune manifestations, more significant respiratory disease and granulomatous changes in pathology; some of these differences can be attributed to co-existence of TACI variants as a genetic modifier. However, numerous other patients with no genetic basis yet discovered, have similar medical histories. It is possible that aside from genetics, these different clinical manifestations result from metabolic, environmental factors or epigenetic causes (Del Pino-Molina et al., 2019; Jorgensen et al., 2019; Macpherson et al., 2019; Ho et al., 2021; Rodriguez-Ubreva et al., 2022; Macpherson et al., 2023). While the majority of patients who carry the "CVID" diagnosis do not yet have a clarified molecular cause, the genetic discoveries in antibody defects continue to reveal the complex immunologic pathways needed to initiate and sustain normal B cell development and the long-term maintenance of B cell memory. With further exploration, more digenic or even polygenic causes of CVID are likely to be dissected, considering the intersecting immunologic pathways.

Data availability statement

The datasets presented in this study are available upon request.

Ethics statement

The studies involving humans were approved by the Mount Sinai School of Medicine. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

CC-R: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing–original draft, Writing–review and editing. J-LC: Data curation, Funding acquisition, Investigation, Methodology, Resources, Software, Supervision, Writing–review and editing. BB: Formal Analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Supervision, Validation, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was partially supported by the NIH NIAID, AI-061093, AI-08603. The Laboratory of Human Genetics of Infectious Diseases was supported by the Howard Hughes Medical Institute, the Rockefeller

University, the St. Giles Foundation, the National Institutes of Health (NIH) (P01AI061093), the French National Research Agency (ANR) under the "Investments for the Future" program (ANR-10-IAHU-01), the French Foundation for Medical Research (FRM; EQU201903007798), the Square Foundation, Grandir—Fonds de solidarité pour l'enfance, the Fondation du Souffle, the SCOR Corporate Foundation for Science, Institut National de la Santé et de la Recherche Médicale (INSERM), and the University of Paris Cité.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1272912/full#supplementary-material

References

Abolhassani, H., Hammarstrom, L., and Cunningham-Rundles, C. (2020). Current genetic landscape in common variable immune deficiency. *Blood* 135 (9), 656–667. doi:10.1182/blood.2019000929

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

Barbosa, R. R., Silva, S. P., Silva, S. L., Tendeiro, R., Melo, A. C., Pedro, E., et al. (2012). Monocyte activation is a feature of common variable immunodeficiency irrespective of plasma lipopolysaccharide levels. *Clin. Exp. Immunol.* 169 (3), 263–272. doi:10.1111/j.1365-2249.2012.04620.x

Bates, C. A., Ellison, M. C., Lynch, D. A., Cool, C. D., Brown, K. K., and Routes, J. M. (2004). Granulomatous-lymphocytic lung disease shortens survival in common variable immunodeficiency. *J. Allergy Clin. Immunol.* 114 (2), 415–421. doi:10.1016/j.jaci.2004.05.057

Bonilla, F. A., Barlan, I., Chapel, H., Costa-Carvalho, B. T., Cunningham-Rundles, C., de la Morena, M. T., et al. (2016). International consensus document (ICON): common variable immunodeficiency disorders. *J. Allergy Clin. Immunol. Pract.* 4 (1), 38–59. doi:10.1016/j.jaip.2015.07.025

Chapel, H., and Cunningham-Rundles, C. (2009). Update in understanding common variable immunodeficiency disorders (CVIDs) and the management of patients with these conditions. *Br. J. Haematol.* 145 (6), 709–727. doi:10.1111/j.1365-2141.2009.07669.x

Chapel, H., Lucas, M., Lee, M., Bjorkander, J., Webster, D., Grimbacher, B., et al. (2008). Common variable immunodeficiency disorders: division into distinct clinical phenotypes. *Blood* 112 (2), 277–286. doi:10.1182/blood-2007-11-124545

Chen, K., Coonrod, E. M., Kumanovics, A., Franks, Z. F., Durtschi, J. D., Margraf, R. L., et al. (2013). Germline mutations in NFKB2 implicate the noncanonical NF-κB pathway in the pathogenesis of common variable immunodeficiency. *Am. J. Hum. Genet.* 93 (5), 812–824. doi:10.1016/j.ajhg.2013.09.009

Cols, M., Rahman, A., Maglione, P. J., Garcia-Carmona, Y., Simchoni, N., Ko, H. M., et al. (2016). Expansion of inflammatory innate lymphoid cells in patients with common variable immune deficiency. *J. Allergy Clin. Immunol.* 137 (4), 1206–1215. doi:10.1016/j.jaci.2015.09.013

Del Pino-Molina, L. R.-U. J., Torres Canizales, J., Coronel-Díaz, M., Kulis, M., Martín-Subero, J. I., van der Burg, M., et al. (2019). Impaired CpG demethylation in common variable immunodeficiency associates with B cell phenotype and proliferation rate. *Front. Immunol.* 10, 878. doi:10.3389/fimmu.2019.00878

Fabre, A., Marchal, S., Barlogis, V., Mari, B., Barbry, P., Rohrlich, P. S., et al. (2019). Clinical aspects of STAT3 gain-of-function germline mutations: a systematic Review. *J. Allergy Clin. Immunol. Pract.* 7 (6), 1958–1969. doi:10.1016/j.jaip.2019.02.018

Fevang, B., Yndestad, A., Sandberg, W. J., Holm, A. M., Muller, F., Aukrust, P., et al. (2007). Low numbers of regulatory T cells in common variable immunodeficiency: association with chronic inflammation *in vivo*. *Clin. Exp. Immunol.* 147 (3), 521–525. doi:10.1111/j.1365-2249.2006.03314.x

Fraz, M. S. A., Michelsen, A. E., Moe, N., Aalokken, T. M., Macpherson, M. E., Nordoy, I., et al. (2022). Raised serum markers of T cell activation and exhaustion in granulomatous-lymphocytic interstitial lung disease in common variable immunodeficiency. *J. Clin. Immunol.* 42 (7), 1553–1563. doi:10.1007/s10875-022-01318-1

Gathmann, B., Mahlaoui, N., Ceredih, Gerard, L., Oksenhendler, E., Warnatz, K., et al. (2014). Clinical picture and treatment of 2212 patients with common variable immunodeficiency. *J. Allergy Clin. Immunol.* 134 (1), 116–126. doi:10.1016/j.jaci.2013.12.1077

Habibi, S., Zaki-Dizaji, M., Rafiemanesh, H., Lo, B., Jamee, M., Gamez-Diaz, L., et al. (2019). Clinical, immunologic, and molecular spectrum of patients with LPS-responsive beige-like anchor protein deficiency: a systematic Review. *J. Allergy Clin. Immunol. Pract.* 7 (7), 2379–2386. doi:10.1016/j.jaip.2019.04.011

Ho, H. E., and Cunningham-Rundles, C. (2022). Seeking relevant biomarkers in common variable immunodeficiency. *Front. Immunol.* 13, 857050. doi:10.3389/fimmu.2022.857050

Ho, H. E., Radigan, L., Bongers, G., El-Shamy, A., and Cunningham-Rundles, C. (2021). Circulating bioactive bacterial DNA is associated with immune activation and complications in common variable immunodeficiency. *JCI Insight* 6 (19), e144777. doi:10.1172/jci.insight.144777

Itan, Y., Shang, L., Boisson, B., Ciancanelli, M. J., Markle, J. G., Martinez-Barricarte, R., et al. (2016). The mutation significance cutoff: gene-level thresholds for variant predictions. *Nat. Methods* 13 (2), 109–110. doi:10.1038/nmeth.3739

Jamee, M., Moniri, S., Zaki-Dizaji, M., Olbrich, P., Yazdani, R., Jadidi-Niaragh, F., et al. (2020). Clinical, immunological, and genetic features in patients with activated PI3Kδ syndrome (APDS): a systematic Review. *Clin. Rev. Allergy Immunol.* 59 (3), 323–333. doi:10.1007/s12016-019-08738-9

Jorgensen, S. F., Fevang, B., and Aukrust, P. (2019). Autoimmunity and inflammation in CVID: a possible crosstalk between immune activation, gut microbiota, and epigenetic modifications. *J. Clin. Immunol.* 39 (1), 30–36. doi:10.1007/s10875-018-0574-z

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 46 (3), 310–315. doi:10.1038/ng.2892

Klemann, C., Camacho-Ordonez, N., Yang, L., Eskandarian, Z., Rojas-Restrepo, J. L., Frede, N., et al. (2019). Clinical and immunological phenotype of patients with primary immunodeficiency due to damaging mutations in NFKB2. *Front. Immunol.* 10, 297. doi:10.3389/fimmu.2019.00297

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., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp352

Li, J., Lei, W. T., Zhang, P., Rapaport, F., Seeleuthner, Y., Lyu, B., et al. (2021). Biochemically deleterious human NFKB1 variants underlie an autosomal dominant form of common variable immunodeficiency. *J. Exp. Med.* 218 (11), e20210566. doi:10.1084/jem.20210566

Litzman, J., Nechvatalova, J., Xu, J., Ticha, O., Vlkova, M., and Hel, Z. (2012). Chronic immune activation in common variable immunodeficiency (CVID) is associated with elevated serum levels of soluble CD14 and CD25 but not endotoxaemia. *Clin. Exp. Immunol.* 170 (3), 321–332. doi:10.1111/j.1365-2249.2012.04655.x

Lorenzini, T., Fliegauf, M., Klammer, N., Frede, N., Proietti, M., Bulashevska, A., et al. (2020). Characterization of the clinical and immunologic phenotype and management of 157 individuals with 56 distinct heterozygous NFKB1 mutations. *J. Allergy Clin. Immunol.* 146 (4), 901–911. doi:10.1016/j.jaci.2019.11.051

Macpherson, M. E., Halvorsen, B., Yndestad, A., Ueland, T., Mollnes, T. E., Berge, R. K., et al. (2019). Impaired HDL function amplifies systemic inflammation in common variable immunodeficiency. *Sci. Rep.* 9 (1), 9427. doi:10.1038/s41598-019-45861-1

Macpherson, M. E., Skarpengland, T., Hov, J. R., Ranheim, T., Vestad, B., Dahl, T. B., et al. (2023). Increased plasma levels of triglyceride-enriched lipoproteins associate with systemic inflammation, lipopolysaccharides, and gut dysbiosis in common variable immunodeficiency. *J. Clin. Immunol.* 43, 1229–1240. doi:10.1007/s10875-023-01475-x

Maffucci, P., Bigio, B., Rapaport, F., Cobat, A., Borghesi, A., Lopez, M., et al. (2019). Blacklisting variants common in private cohorts but not in public databases optimizes human exome analysis. *Proc. Natl. Acad. Sci. U. S. A.* 116 (3), 950–959. doi:10.1073/pnas.1808403116

Maffucci, P., Filion, C. A., Boisson, B., Itan, Y., Shang, L., Casanova, J. L., et al. (2016). Genetic diagnosis using whole exome sequencing in common variable immunodeficiency. *Front. Immunol.* 7, 220. doi:10.3389/fimmu.2016.00220

Malphettes, M., Gerard, L., Carmagnat, M., Mouillot, G., Vince, N., Boutboul, D., et al. (2009). Late-onset combined immune deficiency: a subset of common variable immunodeficiency with severe T cell defect. *Clin. Infect. Dis.* 49 (9), 1329–1338. doi:10.1086/606059

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

Mouillot, G., Carmagnat, M., Gerard, L., Garnier, J. L., Fieschi, C., Vince, N., et al. (2010). B-cell and T-cell phenotypes in CVID patients correlate with the clinical phenotype of the disease. *J. Clin. Immunol.* 30 (5), 746–755. doi:10.1007/s10875-010-9424-3

Ng, P. C., and Henikoff, S. (2003). SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31 (13), 3812–3814. doi:10.1093/nar/gkg509

Odnoletkova, I., Kindle, G., Quinti, I., Grimbacher, B., Knerr, V., Gathmann, B., et al. (2018). The burden of common variable immunodeficiency disorders: a retrospective analysis of the European Society for Immunodeficiency (ESID) registry data. *Orphanet J. Rare Dis.* 13 (1), 201. doi:10.1186/s13023-018-0941-0

Picard, C., Bobby Gaspar, H., Al-Herz, W., Bousfiha, A., Casanova, J. L., Chatila, T., et al. (2018). International union of immunological societies: 2017 primary immunodeficiency diseases committee report on inborn errors of immunity. *J. Clin. Immunol.* 38 (1), 96–128. doi:10.1007/s10875-017-0464-9

Ramirez, N. J., Posadas-Cantera, S., Caballero-Oteyza, A., Camacho-Ordonez, N., and Grimbacher, B. (2021). There is no gene for CVID - novel monogenetic causes for primary antibody deficiency. *Curr. Opin. Immunol.* 72, 176–185. doi:10.1016/j.coi.2021.05.010

Registry, E. (2016). *Working definitions for clinical diagnosis of PID*, 1–25. Available at: https://esidorg/Working-Parties/Registry/Diagnosis-criteria.

Resnick, E. S., Moshier, E. L., Godbold, J. H., and Cunningham-Rundles, C. (2012a). Morbidity and mortality in common variable immune deficiency over 4 decades. *Blood* 119 (7), 1650–1657. doi:10.1182/blood-2011-09-377945

Resnick, E. S., Moshier, E. L., Godbold, J. H., and Cunningham-Rundles, C. (2012b). Morbidity and mortality in common variable immune deficiency over 4 decades. *Blood* 119 (7), 1650–1657. doi:10.1182/blood-2011-09-377945

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

Rodriguez-Ubreva, J., Arutyunyan, A., Bonder, M. J., Del Pino-Molina, L., Clark, S. J., de la Calle-Fabregat, C., et al. (2022). Single-cell Atlas of common variable immunodeficiency shows germinal center-associated epigenetic dysregulation in B-cell responses. *Nat. Commun.* 13 (1), 1779. doi:10.1038/s41467-022-29450-x

Rojas-Restrepo, J., Caballero-Oteyza, A., Huebscher, K., Haberstroh, H., Fliegauf, M., Keller, B., et al. (2021). Establishing the molecular diagnoses in a cohort of 291 patients with predominantly antibody deficiency by targeted next-generation sequencing: experience from a monocentric study. *Front. Immunol.* 12, 786516. doi:10.3389/fimmu.2021.786516

Salzer, U., Bacchelli, C., Buckridge, S., Pan-Hammarstrom, Q., Jennings, S., Lougaris, V., et al. (2009). Relevance of biallelic versus monoallelic TNFRSF13B mutations in distinguishing disease-causing from risk-increasing TNFRSF13B variants in antibody deficiency syndromes. *Blood* 113 (9), 1967–1976. doi:10.1182/blood-2008-02-141937

Salzer, U., Chapel, H. M., Webster, A. D., Pan-Hammarstrom, Q., Schmitt-Graeff, A., Schlesier, M., et al. (2005). Mutations in TNFRSF13B encoding TACI are associated with common variable immunodeficiency in humans. *Nat. Genet.* 37 (8), 820–828. doi:10.1038/ng1600

Salzer, U., and Grimbacher, B. (2021). TACI deficiency - a complex system out of balance. *Curr. Opin. Immunol.* 71, 81–88. doi:10.1016/j.coi.2021.06.004

Sanchez-Ramon, S., Radigan, L., Yu, J. E., Bard, S., and Cunningham-Rundles, C. (2008). *Memory B cells in common variable immunodeficiency: clinical associations and sex differences*. Orlando, Fla: Clinical immunology.

Schwab, C., Gabrysch, A., Olbrich, P., Patino, V., Warnatz, K., Wolff, D., et al. (2018). Phenotype, penetrance, and treatment of 133 cytotoxic T-lymphocyte antigen 4-insufficient subjects. *J. Allergy Clin. Immunol.* 142 (6), 1932–1946. doi:10.1016/j.jaci.2018.02.055

Seidel, M. G., Kindle, G., Gathmann, B., Quinti, I., Buckland, M., van Montfrans, J., et al. (2019). The European society for immunodeficiencies (ESID) registry working definitions for the clinical diagnosis of inborn errors of immunity. *J. Allergy Clin. Immunol. Pract.* 7, 1763–1770. doi:10.1016/j.jaip.2019.02.004

Smith, T., and Cunningham-Rundles, C. (2021). Lymphoid malignancy in common variable immunodeficiency in a single-center cohort. *Eur. J. Haematol.* 107 (5), 503–516. doi:10.1111/ejh.13687

Stenson, P. D., Ball, E. V., Mort, M., Phillips, A. D., Shiel, J. A., Thomas, N. S., et al. (2003). Human gene mutation database (HGMD): 2003 update. *Hum. Mutat.* 21 (6), 577–581. doi:10.1002/humu.10212

Tuijnenburg, P., Lango Allen, H., Burns, S. O., Greene, D., Jansen, M. H., Staples, E., et al. (2018). Loss-of-function nuclear factor κB subunit 1 (NFKB1) variants are the most common monogenic cause of common variable immunodeficiency in Europeans. *J. Allergy Clin. Immunol.* 142, 1285–1296. doi:10.1016/j.jaci.2018.01.039

Warnatz, K., Denz, A., Drager, R., Braun, M., Groth, C., Wolff-Vorbeck, G., et al. (2002). Severe deficiency of switched memory B cells (CD27(+)IgM(-)IgD(-)) in subgroups of patients with common variable immunodeficiency: a new approach to classify a heterogeneous disease. *Blood* 99 (5), 1544–1551. doi:10.1182/blood.v99.5.1544

Wehr, C., Kivioja, T., Schmitt, C., Ferry, B., Witte, T., Eren, E., et al. (2008). The EUROclass trial: defining subgroups in common variable immunodeficiency. *Blood* 111 (1), 77–85. doi:10.1182/blood-2007-06-091744

Zhang, L., Radigan, L., Salzer, U., Behrens, T. W., Grimbacher, B., Diaz, G., et al. (2007). Transmembrane activator and calcium-modulating cyclophilin ligand interactor mutations in common variable immunodeficiency: clinical and immunologic outcomes in heterozygotes. *J. Allergy Clin. Immunol.* 120 (5), 1178–1185. doi:10.1016/j.jaci.2007.10.001

frontiers | Frontiers in Genetics

TYPE Review
PUBLISHED 04 March 2024
DOI 10.3389/fgene.2024.1341272

OPEN ACCESS

EDITED BY
Seik-Soon Khor,
Nanyang Technological University, Singapore

REVIEWED BY
Emanuele Micaglio,
IRCCS San Donato Polyclinic, Italy
Sheng Wang,
University of California, San Francisco, United States

*CORRESPONDENCE
Joo Guan Yeo,
gmsyeo@duke-nus.edu.sg

†These authors have contributed equally to this work

‡These authors share senior authorship

RECEIVED 20 November 2023
ACCEPTED 20 February 2024
PUBLISHED 04 March 2024

CITATION
Yeo NK-W, Lim CK, Yaung KN, Khoo NKH, Arkachaisri T, Albani S and Yeo JG (2024), Genetic interrogation for sequence and copy number variants in systemic lupus erythematosus.
Front. Genet. 15:1341272.
doi: 10.3389/fgene.2024.1341272

Genetic interrogation for sequence and copy number variants in systemic lupus erythematosus

Nicholas Kim-Wah Yeo[1,2†], Che Kang Lim[2,3†], Katherine Nay Yaung[1,2], Nicholas Kim Huat Khoo[1], Thaschawee Arkachaisri[1,2,4‡], Salvatore Albani[1,2,4‡] and Joo Guan Yeo[1,2,4*‡]

[1]Translational Immunology Institute, SingHealth Duke-NUS Academic Medical Centre, Singapore, Singapore, [2]Duke-NUS Medical School, Singapore, Singapore, [3]Department of Clinical Translation Research, Singapore General Hospital, Singapore, Singapore, [4]Rheumatology and Immunology Service, KK Women's and Children's Hospital, Singapore, Singapore

Early-onset systemic lupus erythematosus presents with a more severe disease and is associated with a greater genetic burden, especially in patients from Black, Asian or Hispanic ancestries. Next-generation sequencing techniques, notably whole exome sequencing, have been extensively used in genomic interrogation studies to identify causal disease variants that are increasingly implicated in the development of autoimmunity. This Review discusses the known casual variants of polygenic and monogenic systemic lupus erythematosus and its implications under certain genetic disparities while suggesting an age-based sequencing strategy to aid in clinical diagnostics and patient management for improved patient care.

KEYWORDS

systemic lupus erythematosus, genomics, next-generation sequencing, whole exome sequencing, monogenic, copy number variation

1 Introduction

Systemic lupus erythematosus (SLE, or lupus) is an autoimmune disease characterized by autoantibody formation targeting nucleic components like double-stranded DNA (dsDNA) and RNA (Caielli et al., 2023). The vast spectrum of clinical manifestations ranges from mild skin rashes to widespread destructive multi-organ inflammation, which in some cases, could result in death. The pathogenesis of SLE is complex and multi-factorial (Tsokos, 2011), with genetic and environmental contributions to the disease. It has also been observed that various autoimmune diseases are more common in women (Vinuesa et al., 2023), and in SLE, individuals from Black, Asian or Hispanic ethnicities have an increased disease burden, with patients presenting with a more severe phenotype (Lewis and Jawad, 2017).

SLE can be grouped according to the age of disease onset into adult- and childhood-onset SLE (cSLE); the latter referring to those diagnosed before the age of 18 years and generally presents with greater severity especially in children under 5 years old (Bundhun et al., 2017; Alperin et al., 2018). This early onset of SLE has been associated with an increased genetic burden, highlighting the contribution of one or several risk alleles to

disease (Webb et al., 2011). And within this patient group, around 3%–10% of patients carry a single disease-causing variant (Almlof et al., 2019; Belot et al., 2020; Charras et al., 2023), thus being increasingly recognized and termed as monogenic SLE (Harley and Sawalha, 2022; Vinuesa et al., 2023). Pinpointing the disease-causing variant will contribute greatly to our current knowledge of lupus pathogenesis, and this can be achieved through the use of next-generation sequencing (NGS) techniques (Sanger et al., 1977; Slatko et al., 2018; You et al., 2018; Yaung et al., 2023). As such, a focused strategy is needed together with prioritizing NGS and research efforts towards cSLE patients (Mina and Brunner, 2013).

Knowing that SLE has a strong genetic component to disease (Lewis and Jawad, 2017), multiple susceptibility loci have since been identified, following the advent of genome-wide association studies (GWAS) (Deng and Tsao, 2014). Further diving into genomic studies of SLE through NGS techniques has brought to light the utility of whole exome (WES) and whole genome sequencing (WGS). Our colleagues have also reviewed various technologies that could be employed to elucidate disease mechanisms (Yaung et al., 2023), such as Sanger sequencing (Sanger et al., 1977), single nucleotide polymorphism (SNP) array (You et al., 2018), WES and WGS (Slatko et al., 2018). In this Review, we expound further into the use of NGS techniques, notably WES, across the current genomic landscape of polygenic and monogenic SLE, discussing its potential in reconciling disease risk variants and copy number variations (CNVs) and evaluating the identification of such variants.

2 Next-generation sequencing in SLE

Sequencing technologies have been fundamental for researchers due to their high-throughput capabilities and more recently, their cost-effectiveness (Goodwin et al., 2016). This has allowed for comprehensive genomic studies (i.e., point mutations, small indels, CNVs) and paved the way for multi-omics studies (Levy and Myers, 2016; Lee et al., 2022; Satam et al., 2023; Yaung et al., 2023). In the context of systemic autoimmune diseases like SLE, multiple susceptibility loci identified by GWAS cumulatively contribute risk towards its development but carry a relatively low disease risk individually (Sestak et al., 2011; Wahren-Herlenius and Dorner, 2013).

Several methods have been employed in SLE genomics, including WGS, WES and targeted sequencing (Table 1). Briefly, WGS allows for comprehensive interrogation of the entire human genome and has contributed significantly to the genomic landscape via the 1000 Genomes project since 2010 (Genomes Project et al., 2010; Genomes Project et al., 2015; Sudmant et al., 2015). However, around 85% of disease-related mutations are concentrated in the exome, which constitutes about 2% of the whole genome (Majewski et al., 2011). WES then involves the selection of protein-coding regions (exons) in the genome for sequencing to identify any changes that could impact protein sequences (Ng et al., 2009). This has led to its increased use due to the significant reduction vis-à-vis the starting material, cost and data management (Petersen et al., 2017). In addition, mutations in the exonic region have been shown to be a major contributor to the development of monogenic diseases (Kuhlenbaumer et al., 2011). With the knowledge obtained from the above-mentioned methods, sequencing panels could be generated to target certain regions of interest that harbor pathogenic mutations, hence the utility of targeted sequencing for potential clinical care (Gulilat et al., 2019).

2.1 Polygenic contribution to SLE

Autoimmune diseases have been known to arise from an accumulation of genetic and environmental factors across one's lifetime, as in the case of adult-onset SLE (Goodnow et al., 2005). More than 100 loci associated with SLE have been identified through GWAS (Wang et al., 2021), such as regions in the Human Leukocyte Antigen (*HLA*) locus (Hanscombe et al., 2018), *STAT4* (Remmers et al., 2007; Han et al., 2009), *TNFSF4* (Han et al., 2009), *BANK1* (Kozyrev et al., 2008; Martinez-Bueno et al., 2018), *TNFAIP3* (Graham et al., 2008; Musone et al., 2008; Han et al., 2009), *BLK* (Hom et al., 2008; Han et al., 2009), *IRF5* (Jones et al., 2019), *ETS1* (Han et al., 2009; Yang et al., 2010; Jones et al., 2019), *WDFY4* (Yang et al., 2010) and *TNIP1* (Han et al., 2009; Yang et al., 2010; Jones et al., 2019). However, these variants are unlikely to contribute significantly to SLE pathogenesis individually, unless coupled either with variants in certain regulatory regions or in other genes that maintain immune tolerance (Jones et al., 2019). Importantly, epistatic interaction between genes may contribute in part to the development of complex diseases such as lupus (Hughes et al., 2012; Wei et al., 2014).

It has been recently suggested that polygenic risk scores (PRS) could be utilized to identify and stratify potential SLE patients for early intervention, if needed (Khunsriraksakul et al., 2023). Briefly, GWAS-identified risk variants are statistically compiled to predict disease incidence in a population and risk for developing SLE in individuals (Khunsriraksakul et al., 2022). An association between a high PRS and poorer prognosis in SLE has been observed (Chen et al., 2020; Reid et al., 2020; Sandling et al., 2021), with one study going further to delineate T cell differentiation and innate immunity as the two key axes of SLE association mediated by HLA and interferons (IFNs) respectively (Sandling et al., 2021). The strong involvement of HLA and IFNs has also been described for SLE pathogenesis (Chen et al., 2017; Villarino et al., 2017; Alunno et al., 2019; Crow and Ronnblom, 2019). Despite its utility, PRS has yet to be generalizable beyond the specific population being studied, which further emphasizes the need for larger, diverse and well-represented datasets in order to draw meaningful conclusions (Torkamani et al., 2018). In addition, data generated from GWAS is primarily based on SNP arrays which can be limited by its inability to identify causal variants and ultra-rare mutations, particularly in ethically under-represented populations (Tam et al., 2019). NGS techniques thus provide an answer to interrogating such variants, which might aid in enriching our knowledge of SLE pathogenesis, the clinical diagnosis and management of polygenic SLE together with the potential use of PRS.

2.2 Monogenic contribution to SLE

Single gene defects are part of the diverse heterogenous etiologies for lupus, where about 1%–3% of SLE patients carry a single mutation that leads to disease development (Costa-Reis and

TABLE 1 NGS techniques and respective applications in SLE studies.

Reference(s)	Study population			Gene(s) identified	Change in sequence		Clinical manifestations (if any)
	Cases	Controls	Ethnicity		Nucleotide	Protein	
SNP array							
Han et al. (2009), Nat Genet	1,047	1,205	Han Chinese	*BLK, ETS1, IKZF1, IRF5, RASGRP3, SLC15A4, STAT4, TNFAIP3, TNFSF4, TNIP1, 6q21 7q11.23, 10q11.22, 11q23.3, 16p11.2, 22q11*	-	-	-
Rioux et al. (2009), Proc Natl Acad Sci	643	1,049	British	*HLA-DRA, TNSB-CREB1*	-	-	-
	483	746	American				
	-	672	Swedish				
Fernando et al. (2012), Ann Rheum Dis	464	468	Spanish	*BTNL2-DRA, C6orf27, DPB1-DPB2, MSH5*	-	-	-
	335	247	Filipino	*DRB1-D1A1, HLA-G-HLA-H, HLA-B-MICA, MSH5, DPB1*			
	632	742	British	*BTNL2-DRA, HLA-C-HLA-B, MUC21-PSORS1C1, TNXB-ATF6B*			
Webb et al. (2011), Ann Rheum Dis	1,569	1893	AA	*BANK1, CFB, CTLA4, C8orf13-BLK, FCGR2A, ITGAM, KIAA1542, MBL2, MECP2, MSH5, PDCD1, PTPN22, PXK, STAT4, TNFSF4*	-	-	-
	155	131	Gullah AA				
Hom et al. (2008), NEJM	1,435	3,583	European	*BLK, HLA, IRF5, ITGAM-ITGAX, STAT4*	-	-	-
	793	857	Swedish				
Gateva et al. (2009), Nat Genet	1,129	2,291	American	*ATG5, BANK1, BLK, FCGR2A, HLA-DRB1, HLA-DRB2, IRAK1-MECP2, ITGAM, IRF5, KIAA1542 (PHRF1), PTPN22, PTTG1, PXK, STAT4, TNFAIP3, TNFSF4 (OX40L), UBE2L3*	-	-	-
	834	1,338	Swedish				
Bentham et al. (2015), Nat Genet	4,036	6,959	European	*ARID5B, BANK1, BLK, CD44, CSK, CXorf21, IFIH1, IKZF1, IKZF2, IKZF3, IL10, IL12A, IRF5, IRF7, IRF8, ITGAM, JAZF1, LYST, MHC Class II, MIR146A, PCGR2A, PLD2, PTPN22, RAD51B, SH2B3, SLC15A4, SPRED2, STAT4, TNFAIP3, TNFSF4, TNIP1, TYK2, UBE2L3, UHRF1BP1, WDFY4*	-	-	-
Wang et al. (2021), Nat Commun	6,707	16,047	East Asian	*DSE, HIP1, IKZF1, NEURL4-ACAP1, PLD4, PRKCB, PRRX1-MROH9, TNFRSF13B, TYK2*	-	-	-
	4,576	8,039	European				
Elghzaly et al. (2022), Front Genet	458	769	Egyptian	*DEF6-PPARD, IRF1, IRF5, ITGAM-ITGAX, TYK2, XKR6*	-	-	-

(Continued on following page)

TABLE 1 (*Continued*) NGS techniques and respective applications in SLE studies.

Reference(s)	Study population			Gene(s) identified	Change in sequence		Clinical manifestations (if any)
	Cases	Controls	Ethnicity		Nucleotide	Protein	
Martínez-Bueno et al. (2018), Front Immunol	4,212	4,065	European	*HLA-DRB1, IRF7, PRDM1, SPATA8, TMEM55B*	-	-	-
Yang et al. (2015), Am J Hum Genet	1,656	3,394	East Asian	*ARID5B, CD80, CDKN1B, DRAM1, GPR19, SREBL2, TET3*	-	-	-
Yang et al. (2010), PLoS Genet	3,614	5,684	East Asian	*BANK1, BLK, ETS1, HLA-DRB1-HLA-DQA1, IRF5, STAT4, TNFAIP3, TNFSF4, WDFY4*	-	-	-
Martínez-Bueno et al. (2018), Int J Mol Sci	4,212	4,065	European	*BANK1*	-	-	-
	1761	1,138	AA				
Hanscombe et al. (2018), Hum Mol Genet	4,036	6,959	European	*HLA-DQA, HLA-DQB*	-	-	-
	1,494	5,908	AA				
WES							
Tirosh et al. (2019), Pediatr Rheumol Online J	15	-	Israeli	C1QC (premature stop codon)	c.271G>T	p.G91*	Lupus exacerbation, MAS, sepsis
				MAN2B1	c.192C>A	p.V56M	Dysmorphic features, decreased breath sounds bilaterally, hepatosplenomegaly, malar rash, diffuse abdominal papulosquamous rash and palmar erythema
				SLC7A7	c.943T>C	p.S315P	
				PTEN	c.697C>T	p.R233X	Macrocephaly, developmental delay, pigmented gums and macules of the glans penis; previously reported by Liaw et al., 1997, Nat Genet.
				STAT1 gain-of-function	c.862A>G	p.T288A	Chronic mucocutaneous candidiasis and autoimmunity
Batu et al. (2018), J Rheumatol	7	245	Turkish	*C1QA*	c.622C>T	p.Q208*	-
				C1QC	c.79C>T	p.Q27*	
					c.100G>A	p.G34R	
				C1S	c.1945G>C	p.A649P	
				DNASE1L3	c.289_290delAC	p.T97lefs*2	
				HDAC7	c.163C>T	p.R55W	
Brown et al. (2022), Nature	1	-	European	*TLR7*	c.790T>C	p.Y264H	Inflammatory arthralgias, constitutional symptoms, intermittent hemichorea episodes with hypertensive crisis
	1	-	American		c.82A>G	p.R28G	Neuromyelitis optica in the presence of ANAs and antibodies to aquaporin-4
	1	-	East Asian		c.1521T>G	p.F507L	Malar rash, join pain, Raynaud's phenomenon, alopecia, fever, oral ulcers

(Continued on following page)

TABLE 1 (*Continued*) NGS techniques and respective applications in SLE studies.

Reference(s)	Study population			Gene(s) identified	Change in sequence		Clinical manifestations (if any)
	Cases	Controls	Ethnicity		Nucleotide	Protein	
Lee et al. (2022), Pediatr Rheumatol Online J	184	-	East Asian	*CFHR4*	c.103T>C	-	FTT, fever, arthritis, discoid rashes
				C1S	c.1241G>A	p.R414H	Oral ulcerations, swelling legs, seizure with posterior reversible encephalopathy syndrome
				C2	c.1558C>T	p.R520C	
				DNASE1	c.370G>A	p.E124K	
				DNASE1L3	c.764G>A	p.R225K	Nephrotic-range proteinuria, malar rash, oral ulceration, arthritis
				SLC7A7	c.625 + 1 G>A	-	Twin 1: transient proteinuria, lymphopenia, thrombocytopenia, low C4, positive auto-Ab profile Twin 2: FTT, glomerulonephritis with profound proteinuria
				TREX1	c.292_293 ins A	p.C99M fs	FTT, chilblain-like skin lesions, dystonic posturing with peripheral spasticity
Delgado-Vega et al. (2018), Sci Rep	5	-	Icelandic	*ANKRD50*	-	p.T367M	-
				CHD3		p.A1523T	
				CLC		p.N65K	
				DCLRE1C		p.H283N	
				FAM71E1		p.L7F	
				FAM8A1		p.G234R	
				FAT4		p.P247T	
				FBXL14		p.N102H	
				KIR2DS4		p.I255L	
				KRTAP4-9		p.D18V	
				MPHOSPH8		p.E499K	
				NOTCH1		p.D932N	
				NUP214		p.I765V	
				PABPC3		p.A114T	
				PDHA2		p.R286P	
				SCL25A9		p.G103R	
				TPRA1		p.E300K	
				WDR25		p.R206H	
				XRCCBP1		p.A229V	
Hong et al. (2022), Front Pediatr	1	-	East Asian	*ACP5*	c.420G>A	p.R46Q	Recurrent upper respiratory tract infections and oral thrush throughout life; presented with cutaneous bleeding spots on lower extremities
					c.1152G>T	p.G290V	
				SAMDH1	c.1423G>A	p.R408H	

(Continued on following page)

TABLE 1 (*Continued*) NGS techniques and respective applications in SLE studies.

Reference(s)	Study population			Gene(s) identified	Change in sequence		Clinical manifestations (if any)
	Cases	Controls	Ethnicity		Nucleotide	Protein	
Li et al. (2020), Medicine (Baltimore)	7	-	East Asian	*NRAS*	c.38A>G	p.G13C	Genetic changes observed in Patients 1-4
				PIK3CD	c.3061G>A	p.E1021K	Genetic changes observed in Patients 6 and 7
				TNFAIP3	c.559C>T	p.Q187X	
Demirkaya et al. (2017), Arthritis Rheumatol	4	5	Turkish	*C1R*	c.1332delT	p.P445L fs*11	Systemic inflammation, presence of ANA, malar rash
Tang and Luo. (2022), Arch Rheumatol	4	-	East Asian	*CR2*	C.2804T>C	p.I935T	-
				ITGB3	c.1960G>A	-	
Raupov et al. (2022), Front Pediatr	2	-	Russian	*RNASEL*	c.1880A>G	p.K627R	Twin 1: petechiae, fever, Henoch-Schonlein purpura Twin 2: acute severe leg pain, petechiae
WGS							
Almlof et al. (2019), Hum Genet	71	2,711	Swedish	*C1QC*	C>T	p.R69*	-
				C1S	G>A	p.D631N	
				DNASE1	G>A	p.G127R	
					C>G	p.P154A	
				DNASE1L3	G>A	p.T224M	
				IFIH1	G>A	p.R77W	
					G>A	p.R374C	
				RNASEH2A	A>G	p.K221R	
Almlof et al. (2021), Eur J Hum Genet	71	2,711	Swedish	*ISX*	c.1076G>A	p.R138N	-
				LTB4R2	c.620C>T	p.N169*	
				MAZ	c.1276T>G	p.C368G	
				PPARA	Deletion 80 kb from TSS in intron 6		
				RBM10	Deletion of exons 3-7		
				SMARCA2	Deletion 603 bases downstream		
Targeted Sequencing							
Alghamdi et al. (2021), Gene	100	147	Egyptian	*AIRE*	AIRE (rs2075876) variant conferred protection against developing SLE, but not the CTLA4 (rs231775) variant		
				CTLA-4			
Lundtoft et al. (2022), Arthritis Rheumatol	2,290	1,251	Scandinavian	*C4*	Low copy number associated with increased risk for SLE (stronger for C4A than C4B)		
Montufar-Robles et al. (2019), Cell Immunol	379	460	Mexican	*AIRE*	Ser196Ser synonymous variant associated with SLE		
Mueller et al. (2013), Am J Hum Genet	4	-	West African	*FCGR3B*	Reduced copy number associated with SLE		
	3		East Asian				
	2		CEPH				
Raj et al. (2020), Genome Biol	1700	2,108	Caucasian	*DAP1*	Risk haplotype downregulates transcription, leading to increased autophagy, diminished apoptosis, increased humoral autoimmunity, increased risk for discoid rash		

(Continued on following page)

TABLE 1 (*Continued*) NGS techniques and respective applications in SLE studies.

Reference(s)	Study population			Gene(s) identified	Change in sequence		Clinical manifestations (if any)
	Cases	Controls	Ethnicity		Nucleotide	Protein	
Sandling et al. (2021), Ann Rheum Dis	958	1,026	Swedish	*ARRDC5*	-	p.N231R	-
				BNC2		p.H307Y	
				C1orf27		p.N246K	
				CAD		p.M922K	
				CAMK2G		p.H370Q	
				CCDC141		p.S1522F	
				DAD1		p.S66R	
				IFNA21		p.Y146C	
				IQGAP1		p.V1371M	
				KIR3DL3		p.G219D	
				MUC5B		p.T2727P	
				PIK3R2		p.G372S	
				PMEL		p.P402H	
				RUNDC1		p.V477I	
				SCNN1A		p.R511*	
				SELENBP1		p.P36T	
				SENP1		p.T155A	
				TMEM132C		p.P1094A	

AA, African-American; ANA, anti-nuclear antibody; CEPH: Centre d'Etude du Polymorphisme Humain; del, deletion; fs, frameshift; FTT, failure-to-thrive; ins, insertion; MAS, macrophage activation syndrome; NGS, next generation sequencing; SLE, systemic lupus erythematosus; SNP, single nucleotide polymorphism; WES, whole exome sequencing; WGS, whole genome sequencing. Italic values denote genes.

Sullivan, 2017). Albeit rare, monogenic SLE is characterized by a more severe phenotype early in life (Webb et al., 2011; Vinuesa et al., 2023). Several gene sets involved in the complement pathway, IFN responses, nucleic acid sensing and immune tolerance have been implicated in the pathogenesis of monogenic SLE (Alperin et al., 2018; Vinuesa et al., 2023).

Most complement-related SLE defects are found in the C1 (*C1QA*, *C1QB*, *C1QC*, *C1R*, *C1S)* (Lood et al., 2009; Bienaime et al., 2010; Demirkaya et al., 2017; Almlof et al., 2019) or C4 (*C4A*, *C4B*) compartments (Blanchong et al., 2001; Vinuesa et al., 2023). Other affected regions include the *C2* and *C3* genes (Miller and Atkinson, 2012; Vinuesa et al., 2023). Complement-deficient patients tend to have an impaired clearance of cellular apoptotic fragments, which in turn facilitates autoimmunity (Costa-Reis and Sullivan, 2017; Vinuesa et al., 2023). Next, studies have observed an elevated IFN signature in SLE patients (Baechler et al., 2003; Reynier et al., 2011), and the association of variants in *ADAR1* (Crow and Ronnblom, 2019), *TREX1* (Rice et al., 2015), *SAMHD1* (Abdel-Salam et al., 2010; Ravenscroft et al., 2011) and *IFIH1* (Rice et al., 2014; Almlof et al., 2019) genes to disease (Abdel-Salam et al., 2010; Ravenscroft et al., 2011; Rice et al., 2012; Rice et al., 2014; Crow et al., 2015). *ADAR1*, *TREX1* and *SAMHD1* are involved in nucleic acid metabolism while *IFIH1* is involved in nucleic acid sensing. Lastly, defects in nucleic acid sensing and degradation genes *DNASE1* (Yasutomo et al., 2001; Almlof et al., 2019), *DNASE1L3* (Al-Mayouf et al., 2011; Almlof et al., 2019) and *TLR7* (Giltiay et al., 2013; Brown et al., 2022) have been found in SLE patients. This causes an accumulation of extracellular nucleic acids, leading to TLR7 activation and downstream type I IFN production. Type I IFN further upregulates TLR7 expression, creating a positive feedback loop that eventuates into autoantibody production (Caielli et al., 2023). With this, it has been recently found that two *TLR7* variants (Y264H and F507L) were recently identified to cause SLE, with the Y264H variant presenting with an increased sensing of guanosine and 2′,3′-cGMP (Shibata et al., 2016; Zhang et al., 2016; Zhang et al., 2018; Brown et al., 2022).

Across autoimmune diseases, a hallmark of its development is the loss of tolerance to self-antigens, with *AIRE* and *CTLA-4* being implicated in SLE (Pullmann et al., 1999; Ahmed et al., 2001; Hudson et al., 2002; Lee et al., 2005; Cunninghame Graham et al., 2006; Lovewell et al., 2015; Montufar-Robles et al., 2019; Alghamdi et al., 2021). *AIRE*, or autoimmune regulator is essential for maintaining central immune tolerance by controlling the negative thymic selection of hyper-reactive T lymphocytes against self-antigens (Yang et al., 2015). Mutations in this gene region have been observed in Norwegian patients with autoimmune polyendocrine syndrome type I (APS-1; (Oftedal et al., 2023)) and Japanese patients with rheumatoid arthritis (RA; (Terao

et al., 2011)). More recently, an *AIRE* Ser196Ser synonymous variant was associated with SLE in a recent targeted sequencing study in a Mexican cohort (Montufar-Robles et al., 2019). However, when extended to GWAS performed on a larger European SLE cohort, no association was found (Bentham et al., 2015). Next, cytotoxic T-lymphocyte associated protein 4 (CTLA-4, or CD152) is an important checkpoint inhibitor in peripheral immune tolerance via negative signaling in regulating autoreactive T cells (Liu and Zhang, 2013; Van Coillie et al., 2020). Though several reports have identified certain polymorphisms contributing to SLE development (Pullmann et al., 1999; Ahmed et al., 2001; Hudson et al., 2002; Lee et al., 2005; Cunninghame Graham et al., 2006; Jury et al., 2010), a meta-analysis has highlighted no association of said variants to lupus (Liu and Zhang, 2013; Alghamdi et al., 2021). In some cases, specific *CTLA-4* variants could even contribute to protection against SLE (Barreto et al., 2004), suggesting that only certain variants within the *CTLA-4* gene region have an association with SLE development.

Recent studies have described several novel genes associated with SLE following WES analysis in an Asian population, such as the decreased expression of cell division cycle 27 (CDC27) in patients (Shang et al., 2022), and novel variants in genes encoding for complement receptor 2 (CR2) (Tang and Luo, 2022), C1R (Demirkaya et al., 2017), NRAS, TNFAIP3 and PIK3CD (Li et al., 2020), WNT16 and ERVW-1 (Chen et al., 2022), ACP5 and SAMHD1 (Hong et al., 2022). This list of genes contributing to monogenic SLE continues to grow with increased usage of WES over the past few years, further enriching our knowledge about the genetic contribution to SLE.

2.3 Copy number variation (CNV)

Copy number variation (CNV) is a phenomenon where repeated genomic sequences occur and arise from the process of genomic rearrangement, which can manifest as translocations, inversions, insertions and deletions (Feuk et al., 2006; Human Genome Structural Variation Working et al., 2007). However, the total number of gene copies and its downstream effects may vary between individuals (Usher and McCarroll, 2015). In the past two decades, several CNVs associated with SLE development have been identified, such as C4 (*C4A*, *C4B*) (Yang et al., 2007; Pereira et al., 2019; Kamitaki et al., 2020; Lundtoft et al., 2022), *FCGR3A*, *FCGR3B* (Willcocks et al., 2008; Niederer et al., 2010), *CCL3L1* (Gonzalez et al., 2005), *RABGAP1L* (Kim et al., 2013), *TLR7* (Garcia-Ortiz et al., 2010) and *HSP90* (Zhang et al., 2019).

As previously mentioned, defects in complement genes have been observed to be a monogenic cause of SLE. Of note, C4, or complement compartment protein 4, is usually present in most individuals as two copies of *C4A* and *C4B* respectively. In some cases, SLE patients may carry a range of zero to five copies of *C4A* and zero to four copies of *C4B* (Yang et al., 2007; Pereira et al., 2019). A recent study has described an association between a low C4A copy number and an increased risk of developing SLE (Kamitaki et al., 2020). Though C4 genes are highly homologous and are usually excluded from variant calling analysis, Lundtoft *et al.* performed a focused analysis into C4 CNVs via targeted sequencing and found Scandinavian SLE patients with a low *C4A* copy number and carrying a common loss-of-function (LoF) variant presenting with lowered plasma C4 levels (Lundtoft et al., 2022). Whether this phenomenon can be extended to other ancestral populations remains unknown and warrants further investigation.

Other genes like *FCG3RA* and *FCGR3B* encode for low-affinity Fc gamma (Fcγ) receptors of IgG and are crucial in the binding and clearing of immune complexes (Willcocks et al., 2008; Niederer et al., 2010), while *CCL3L1* (C-C chemokine ligand 3 like-1) translates into a ligand that binds to C-C chemokine receptor 5 (CCR5) (Gonzalez et al., 2005). Healthy individuals carry two copies of each respective gene, but SLE risk increases when there are either lower or higher copy numbers of said genes (Willcocks et al., 2008). Increased SLE susceptibility was also observed with low *RASGAP1L* and high *TLR7* copy numbers respectively. *RASGAP1L* encodes for a Rab GTPase-activating protein (Kim et al., 2013), while *TLR7* is a key receptor in innate immunity that recognizes single-stranded RNA (Lund et al., 2004; Takeda and Akira, 2005). Lastly, abnormal CNVs in heat shock proteins 90 (HSP90), especially in its AB1 isoform, were identified to correlate with SLE in the Han Chinese (Zhang et al., 2019). This highlights the importance of CNVs in SLE and autoimmunity and thus the need for more traction toward implementing a pipeline to include them in future genetic screens (Zhao et al., 2020).

2.4 Identification of potential disease-causing variants

Genetic testing using NGS techniques has identified potential disease-causing variants and led to better preventative risk management of diseases (Shaw et al., 2023). However, given its complexity, the labeling of variants as potentially pathogenic should be done with caution to prevent misdiagnoses. A misdiagnosis of a pathogenic variant can result in unnecessary medical interventions and cause undue psychological distress to both patients and their families (Manrai et al., 2016; Shaw et al., 2023). Such detrimental consequences have occurred in diseases like hypertrophic cardiomyopathy and cancers, where variants that were thought to be pathogenic were subsequently found to be benign due to the under-representation of certain ancestries in reference control groups (Manrai et al., 2016; Shaw et al., 2023).

To prevent such genetic misclassifications, the American College of Medical Genetics and Genomics has introduced a standardized framework for variant interpretation (Richards et al., 2015). In the case of SLE and other autoimmune diseases, pathogenic variants can be better identified prior to further functional validation through this framework, thus reducing the occurrence of false positives as the number of sequencing studies continues to rise (Vinuesa et al., 2023). In addition, various consortia like the Clinical Genome Resource (ClinGen; (Rehm et al., 2015)), Rheumatologic Autoimmune Clinical Domain Working Group under ClinGen, Lupus in Minority Populations, Nature *versus* Nurture (LUMINA; (Alarcon et al., 2001)) have been established to aggregate all available genomic data and concentrate global research efforts. Crucially, the consolidation of genomic data overcomes the major limitation of genome-wide studies of requiring large sample sizes due to the need to adopt a high level of significance to account for multiple testing (Tam et al., 2019).

With this framework for variant interpretation and genomic data from various consortia, this can be potentially applied to the dysmorphic syndromes associated with SLE, specifically genes of the Ras/mitogen-activated protein kinase (Ras/MAPK) pathway to identify with greater certainty the potential pathogenic genetic variants within this pathway that contribute to SLE (Amoroso et al., 2003; Lisbona et al., 2009; Leventopoulos et al., 2010; Hanaya et al., 2017; Uehara et al., 2018). However, further investigations would be needed to delineate the underlying mechanism with functional studies of the different genes in the Ras/MAPK pathway as these are currently described in case reports and series.

Our current understanding has informed us that certain ancestral groups have an increased predilection towards developing SLE (Lewis and Jawad, 2017), which requires controlling for in future sequencing studies to prevent any potential misclassification of disease-causing variants due to the unavailability of an adequate ancestry-specific reference genome. Past research has been largely focused on European ancestry (Yang et al., 2007; Lewis and Jawad, 2017; Hanscombe et al., 2018), resulting in an under-representation of data from other ancestries to draw meaningful generalizations about the disease. This can be resolved by tapping on several biobanks that have been consolidated over the years to provide greater depth and insights into the genetic differences within and across various ancestries. These include, and are not limited to, the Tohoku Biobank (150,000 participants; (Minegishi et al., 2019)), Mexican Biobank (6,057 participants; (Sohail et al., 2023)), Biobank Japan (BBJ, 260,000 participants; (Kanai et al., 2018)), China Kadoorie Biobank (500,000 participants; (Chen et al., 2011)), H3Africa (70,000 participants; (Consortium et al., 2014; Mulder et al., 2018)), UK Biobank (500,000 participants; (Bycroft et al., 2018; Van Hout et al., 2020; Gaynor et al., 2023)), Michigan Genomic Initiative (MGI, 91,000 participants; (Zawistowski et al., 2023)), Vanderbilt University Biobank (BioVU, 300,000 participants; (Khunsriraksakul et al., 2023)), and SG10K (9,051 participants; (Chan et al., 2022)). It should be noted that SG10K has since been expanded to SG100K, whereby data from 70,000 participants across four national cohort studies will be pooled together with the additional recruitment of 30,000 individuals (Begum, 2022).

3 Discussion

In this Review, we have provided an overview of various susceptibility genes contributing to the development of SLE either through a polygenic or monogenic route identified via NGS techniques, highlighted the involvement and importance of CNVs and urged for the inclusiveness of control groups to account for ancestral differences to prevent any potential variant misclassification.

The introduction of WGS and WES has resulted in faster genomic interrogation, allowing for one's entire genome to be generated in a matter of days to weeks (Bourchany et al., 2017; Duncavage et al., 2021). The data generated from WGS provides comprehensive information on both intronic (non-coding) and exonic (protein-coding) regions. However, the contributions of non-coding variants towards disease have yet to be thoroughly elucidated and the downstream analyses of such intronic regions remain complex and highly challenging (Zhao et al., 2020). As such, WES has become increasingly popular in clinical diagnostics and research due to its utility (~95% capture of exonic and splice site regions (Field, 2021; Zhang et al., 2021)), ease of analysis (Yaung et al., 2023) and lower cost (one-third that of WGS (Goodwin et al., 2016; Field, 2021))

In addition, structural variants like CNV are relatively common across the whole genome (with a frequency of around 12%; (Iafrate et al., 2004; Sebat et al., 2004; Tuzun et al., 2005; Conrad et al., 2006; McCarroll et al., 2006)), and can influence gene expression (Somerville et al., 2005; Lee et al., 2006; McCarroll et al., 2006). As we have alluded to the growing importance of CNV in SLE immunogenetics, the coupling of WES with CNV detection addresses the need for a holistic interrogation of the genetic contribution to SLE through the dual identification of variations in exonic sequences and gene copy numbers. This is achievable with tools such as CoNIFER (Krumm et al., 2012), exomeCopy (Love et al., 2011), CNVkit (Talevich et al., 2016), cn. MOPS (Klambauer et al., 2012), CNest (Fitzgerald and Birney, 2022), CNVind (Kusmirek and Nowak, 2022), CoverageMaster (Rapti et al., 2022) and EXCAVATOR2 (D'Aurizio et al., 2016; D'Aurizio et al., 2018). More recently, Olfe *et al.* have demonstrated CTLA-4 insufficiency due to a novel *CTLA-4* deletion using ClinCNV (German and Stephan, 2019; Olfe et al., 2023), further highlighting the synergy of CNV calling with WES analysis. Beyond the scope of autoimmune diseases, NGS techniques have also been extensively utilized in identifying causal variants (including CNVs) contributing to cancer (van Dijk et al., 2014; Papp et al., 2021; Satam et al., 2023), congenital (Lai et al., 2021; Li et al., 2022; Liu et al., 2022; Wang et al., 2022; Wu et al., 2022; Refeat et al., 2023), cardiovascular (Hu et al., 2023) and hematological diseases (Hassan et al., 2023).

Though the method of WES has been well-established over the years, notable limitations persist in WES-based CNV analyses. The technique primarily targets coding regions, leading to a restricted view of the genome and potentially missing important regulatory components within non-coding regions such as intergenic or intronic regions (Mandelker et al., 2016; Royer-Bertrand et al., 2021). This significantly impacts the sensitivity of CNV detection. In addition, it is susceptible to biases, such as GC content bias, which can impact the reliability of CNV calls (Lelieveld et al., 2015). Furthermore, a relatively higher false positive rate and the limitation of achieving homogeneous coverage of sequencing reads restrict its inclusion as a gold-standard method for CNV detection (Marchuk et al., 2018; Burdick et al., 2020). These limitations emphasize the necessity of integrating WES with other omics approaches for more accuracy in CNV detection (Gabrielaite et al., 2021). Nonetheless, with ongoing upgrades to sequencing libraries, capture kits and bioinformatics pipelines, it is anticipated that the existing limitations will be alleviated (Zhou et al., 2021). Future applications of third-generation sequencing (TGS) techniques such as long-read sequencing hold promise in addressing these constraints and provide additional possibilities in detecting structural variations (SVs) (Xiao and Zhou, 2020).

Though SLE is known to have a strong genetic predilection, its typical development is usually due to polygenic contributions

coupled with an environmental trigger (Harley and Sawalha, 2022); the latter of which must not be ignored. Research into the host-environment interplay has yielded physical/chemical factors (smoking, chemical exposure (Kilburn and Warshaw, 1992; Speyer and Costenbader, 2018; Akhil et al., 2023)), Epstein-Barr virus (EBV) infections (Poole et al., 2006; Jog and James, 2020), gut microbiota (Neuman and Koren, 2017) and obesity (Kang et al., 2020) as contributors to the development of SLE (Parks et al., 2017; Gulati and Brunner, 2018; Akhil et al., 2023). Such environmental triggers can influence methylation patterns in genes related to B and T cells, which are associated with SLE pathology (Akhil et al., 2023). These include observations of hypomethylation in *CD40L* (Vordenbaumen et al., 2021) and *CD70* (Keshavarz-Fathi et al., 2022), as well as hypermethylation of *FOXP3* (Hanaei et al., 2020) and *CTLA-4* (Nosrat zehi et al., 2021).

Elucidating the pathogenesis of autoimmune diseases like SLE remains complex, and studies have called for the need for a multi-omics approach to furnish our current understanding of the disease (Fang et al., 2016; Hedrich, 2017; Kwon et al., 2019; Yaung et al., 2023). Thus far, transcriptomic signatures obtained from blood and tissues have shown an enrichment of genes involved in the IFN response (Banchereau et al., 2016; Der et al., 2019), which corroborates with previous genetic data (Baechler et al., 2003; Reynier et al., 2011). Epigenetic modifications in the genome such as methylation (Ballestar, 2011; Hedrich, 2017), non-coding RNAs (Taheri et al., 2020) and post-translational histone modifications (i.e., methylation, acetylation; (Hu et al., 2008)) have also been associated with the development of SLE. Proteomic studies have proven difficult to isolate biomarkers for diagnosis, management and monitoring due to the heterogeneity of the disease and its involvement across multiple organs (ref), but current efforts continue to show some promise (Huang et al., 2022; Fasano et al., 2023). Indeed, more needs to be done to reconcile multi-omics and genetic data of SLE in the future.

4 Conclusion

Up to 10% of patients below the age of 18 years can carry a significant disease-causing variant which manifests as severe SLE, alluding to a monogenic etiology and highlights the value of doing NGS in children with a very early onset of disease (Alperin et al., 2018; Charras et al., 2021). Previous studies have shown the utility of WES in unraveling novel rare variants and determining its respective contribution(s) to disease (Pullabhatla et al., 2018; Almlof et al., 2019; Tirosh et al., 2019; Almlof et al., 2021). However, genetic variation across ancestries should not be overlooked to prevent variant misclassification and downstream misdiagnoses. This can be controlled via the inclusion of gene datasets across various biobanks, consortia and databases. With that, establishing a pipeline where WES and CNV detection are coupled together will allow for the timely and pinpoint clinical diagnosis of SLE to allow for better clinical management and intervention.

5 Search strategy and selection criteria

We searched PubMed between 30 August 2023 and 7 February 2024, using the terms "systemic lupus erythematosus (SLE)", "next-generation sequencing (NGS)", "genomics", "copy number variation" in articles published from 1 Jan 2013 until 7 February 2024. Articles were also identified through references from articles identified through the search. Only papers published in English were reviewed and the final reference list was generated based on the relevance to the scope of this Review.

Author contributions

NK-WY: Data curation, Writing–original draft, Writing–review and editing. CL: Data curation, Writing–original draft, Writing–review and editing. KN: Writing–review and editing. NK: Writing–review and editing. TA: Writing–review and editing. SA: Writing–review and editing. JY: Conceptualization, Data curation, Supervision, Writing–original draft, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research is supported by the Singapore Ministry of Health's National Medical Research Council (NMRC) under its Centre Grant Programme (MOH-000988) and other NMRC grants: CSAINV22jul-0008 (JY), CIRG21nov-0031 (JY), NMRC/MOHIAFCAT2/005/2015 (SA), NMRC/TCR/0015-NCC/2016 (SA), NMRC/OFLCG/002/2018 (SA) and CIRG19may0052 (SA), is gratefully acknowledged. We also gratefully acknowledge the SingHealth Duke-NUS Academic Medicine Grant, Special Category (PRISM, AM/PRM002/2018).

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 author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Abdel-Salam, G. M., El-Kamah, G. Y., Rice, G. I., El-Darouti, M., Gornall, H., Szynkiewicz, M., et al. (2010). Chilblains as a diagnostic sign of aicardi-goutieres syndrome. *Neuropediatrics* 41 (1), 18–23. doi:10.1055/s-0030-1255059

Ahmed, S., Ihara, K., Kanemitsu, S., Nakashima, H., Otsuka, T., Tsuzaka, K., et al. (2001). Association of CTLA-4 but not CD28 gene polymorphisms with systemic lupus erythematosus in the Japanese population. *Rheumatol. Oxf.* 40 (6), 662–667. doi:10.1093/rheumatology/40.6.662

Akhil, A., Bansal, R., Anupam, K., Tandon, A., and Bhatnagar, A. (2023). Systemic lupus erythematosus: latest insight into etiopathogenesis. *Rheumatol. Int.* 43 (8), 1381–1393. doi:10.1007/s00296-023-05346-x

Alarcon, G. S., McGwin, G., Jr., Bartolucci, A. A., Roseman, J., Lisse, J., Fessler, B. J., et al. (2001). Systemic lupus erythematosus in three ethnic groups. IX. Differences in damage accrual. *Arthritis Rheum.* 44 (12), 2797–2806. doi:10.1002/1529-0131(200112)44:12<2797::aid-art467>3.0.co;2-9

Alghamdi, S. A., Kattan, S. W., Toraih, E. A., Alrowaili, M. G., Fawzy, M. S., and Elshazli, R. M. (2021). Association of AIRE (rs2075876), but not CTLA4 (rs231775) polymorphisms with systemic lupus erythematosus. *Gene* 768, 145270. doi:10.1016/j.gene.2020.145270

Al-Mayouf, S. M., Sunker, A., Abdwani, R., Abrawi, S. A., Almurshedi, F., Alhashmi, N., et al. (2011). Loss-of-function variant in DNASE1L3 causes a familial form of systemic lupus erythematosus. *Nat. Genet.* 43 (12), 1186–1188. doi:10.1038/ng.975

Almlof, J. C., Nystedt, S., Leonard, D., Eloranta, M. L., Grosso, G., Sjowall, C., et al. (2019). Whole-genome sequencing identifies complex contributions to genetic risk by variants in genes causing monogenic systemic lupus erythematosus. *Hum. Genet.* 138 (2), 141–150. doi:10.1007/s00439-018-01966-7

Almlof, J. C., Nystedt, S., Mechtidou, A., Leonard, D., Eloranta, M. L., Grosso, G., et al. (2021). Contributions of *de novo* variants to systemic lupus erythematosus. *Eur. J. Hum. Genet.* 29 (1), 184–193. doi:10.1038/s41431-020-0698-5

Alperin, J. M., Ortiz-Fernandez, L., and Sawalha, A. H. (2018). Monogenic lupus: a developing paradigm of disease. *Front. Immunol.* 9, 2496. doi:10.3389/fimmu.2018.02496

Alunno, A., Padjen, I., Fanouriakis, A., and Boumpas, D. T. (2019). Pathogenic and therapeutic relevance of JAK/STAT signaling in systemic lupus erythematosus: integration of distinct inflammatory pathways and the prospect of their inhibition with an oral agent. *Cells* 8 (8), 898. doi:10.3390/cells8080898

Amoroso, A., Garzia, P., Vadacca, M., Galluzzo, S., Del Porto, F., Mitterhofer, A. P., et al. (2003). The unusual association of three autoimmune diseases in a patient with Noonan syndrome. *J. Adolesc. Health* 32 (1), 94–97. doi:10.1016/s1054-139x(02)00364-6

Baechler, E. C., Batliwalla, F. M., Karypis, G., Gaffney, P. M., Ortmann, W. A., Espe, K. J., et al. (2003). Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc. Natl. Acad. Sci. U. S. A.* 100 (5), 2610–2615. doi:10.1073/pnas.0337679100

Ballestar, E. (2011). Epigenetic alterations in autoimmune rheumatic diseases. *Nat. Rev. Rheumatol.* 7 (5), 263–271. doi:10.1038/nrrheum.2011.16

Banchereau, R., Hong, S., Cantarel, B., Baldwin, N., Baisch, J., Edens, M., et al. (2016). Personalized immunomonitoring uncovers molecular networks that stratify lupus patients. *Cell* 165 (6), 1548–1550. doi:10.1016/j.cell.2016.05.057

Barreto, M., Santos, E., Ferreira, R., Fesel, C., Fontes, M. F., Pereira, C., et al. (2004). Evidence for CTLA4 as a susceptibility gene for systemic lupus erythematosus. *Eur. J. Hum. Genet.* 12 (8), 620–626. doi:10.1038/sj.ejhg.5201214

Batu, E. D., Koşukcu, C., Taşkıran, E., Sahin, S., Akman, S., Sözeri, B., et al. (2018). Whole exome sequencing in early-onset systemic lupus erythematosus. *J. Rheumatol.* 45 (12), 1671–1679. doi:10.3899/jrheum.171358

Begum, S. (2022). Project SG100K: DNA of 100,000 Singaporeans to be mapped to identify new ways to prevent diseases. *Straits Times*.

Belot, A., Rice, G. I., Omarjee, S. O., Rouchon, Q., Smith, E. M. D., Moreews, M., et al. (2020). Contribution of rare and predicted pathogenic gene variants to childhood-onset lupus: a large, genetic panel analysis of British and French cohorts. *Lancet Rheumatol.* 2 (2), e99–e109. doi:10.1016/S2665-9913(19)30142-0

Bentham, J., Morris, D. L., Graham, D. S. C., 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

Bienaime, F., Quartier, P., Dragon-Durey, M. A., Fremeaux-Bacchi, V., Bader-Meunier, B., Patey, N., et al. (2010). Lupus nephritis associated with complete C1s deficiency efficiently treated with rituximab: a case report. *Arthritis Care Res. Hob.* 62 (9), 1346–1350. doi:10.1002/acr.20163

Blanchong, C. A., Chung, E. K., Rupert, K. L., Yang, Y., Yang, Z., Zhou, B., et al. (2001). Genetic, structural and functional diversities of human complement components C4A and C4B and their mouse homologues, Slp and C4. *Int. Immunopharmacol.* 1 (3), 365–392. doi:10.1016/s1567-5769(01)00019-4

Bourchany, A., Thauvin-Robinet, C., Lehalle, D., Bruel, A. L., Masurel-Paulet, A., Jean, N., et al. (2017). Reducing diagnostic turnaround times of exome sequencing for families requiring timely diagnoses. *Eur. J. Med. Genet.* 60 (11), 595–604. doi:10.1016/j.ejmg.2017.08.011

Brown, G. J., Canete, P. F., Wang, H., Medhavy, A., Bones, J., Roco, J. A., et al. (2022). TLR7 gain-of-function genetic variation causes human lupus. *Nature* 605 (7909), 349–356. doi:10.1038/s41586-022-04642-z

Bundhun, P. K., Soogund, M. Z., and Huang, F. (2017). Impact of systemic lupus erythematosus on maternal and fetal outcomes following pregnancy: a meta-analysis of studies published between years 2001-2016. *J. Autoimmun.* 79, 17–27. doi:10.1016/j.jaut.2017.02.009

Burdick, K. J., Cogan, J. D., Rives, L. C., Robertson, A. K., Koziura, M. E., Brokamp, E., et al. (2020). Limitations of exome sequencing in detecting rare and undiagnosed diseases. *Am. J. Med. Genet. A* 182 (6), 1400–1406. doi:10.1002/ajmg.a.61558

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562 (7726), 203–209. doi:10.1038/s41586-018-0579-z

Caielli, S., Wan, Z., and Pascual, V. (2023). Systemic lupus erythematosus pathogenesis: interferon and beyond. *Annu. Rev. Immunol.* 41, 533–560. doi:10.1146/annurev-immunol-101921-042422

Chan, S. H., Bylstra, Y., Teo, J. X., Kuan, J. L., Bertin, N., Gonzalez-Porta, M., et al. (2022). Analysis of clinically relevant variants from ancestrally diverse Asian genomes. *Nat. Commun.* 13 (1), 6694. doi:10.1038/s41467-022-34116-9

Charras, A., Haldenby, S., Smith, E. M. D., Egbivwie, N., Olohan, L., Kenny, J. G., et al. (2023). Panel sequencing links rare, likely damaging gene variants with distinct clinical phenotypes and outcomes in juvenile-onset SLE. *Rheumatol. Oxf.* 62 (SI2), SI210–SI225. doi:10.1093/rheumatology/keac275

Charras, A., Smith, E., and Hedrich, C. M. (2021). Systemic lupus erythematosus in children and young people. *Curr. Rheumatol. Rep.* 23 (3), 20. doi:10.1007/s11926-021-00985-0

Chen, J., Zhang, P., Chen, H., Wang, X., He, X., Zhong, J., et al. (2022). Whole-genome sequencing identifies rare missense variants of WNT16 and ERVW-1 causing the systemic lupus erythematosus. *Genomics* 114 (3), 110332. doi:10.1016/j.ygeno.2022.110332

Chen, L., Morris, D. L., and Vyse, T. J. (2017). Genetic advances in systemic lupus erythematosus: an update. *Curr. Opin. Rheumatol.* 29 (5), 423–433. doi:10.1097/BOR.0000000000000411

Chen, L., Wang, Y. F., Liu, L., Bielowka, A., Ahmed, R., Zhang, H., et al. (2020). Genome-wide assessment of genetic risk for systemic lupus erythematosus and disease severity. *Hum. Mol. Genet.* 29 (10), 1745–1756. doi:10.1093/hmg/ddaa030

Chen, Z., Chen, J., Collins, R., Guo, Y., Peto, R., Wu, F., et al. (2011). China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int. J. Epidemiol.* 40 (6), 1652–1666. doi:10.1093/ije/dyr120

Conrad, D. F., Andrews, T. D., Carter, N. P., Hurles, M. E., and Pritchard, J. K. (2006). A high-resolution survey of deletion polymorphism in the human genome. *Nat. Genet.* 38 (1), 75–81. doi:10.1038/ng1697

Consortium, H. A., Rotimi, C., Abayomi, A., Abimiku, A., Adabayeri, V. M., Adebamowo, C., et al. (2014). Research capacity. Enabling the genomic revolution in Africa. *Science* 344 (6190), 1346–1348. doi:10.1126/science.1251546

Costa-Reis, P., and Sullivan, K. E. (2017). Monogenic lupus: it's all new. *Curr. Opin. Immunol.* 49, 87–95. doi:10.1016/j.coi.2017.10.008

Crow, M. K., and Ronnblom, L. (2019). Type I interferons in host defence and inflammatory diseases. *Lupus Sci. Med.* 6 (1), e000336. doi:10.1136/lupus-2019-000336

Crow, Y. J., Chase, D. S., Lowenstein Schmidt, J., Szynkiewicz, M., Forte, G. M., Gornall, H. L., et al. (2015). Characterization of human disease phenotypes associated with mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR, and IFIH1. *Am. J. Med. Genet. A* 167A (2), 296–312. doi:10.1002/ajmg.a.36887

Cunninghame Graham, D. S., Wong, A. K., McHugh, N. J., Whittaker, J. C., and Vyse, T. J. (2006). Evidence for unique association signals in SLE at the CD28-CTLA4-ICOS locus in a family-based study. *Hum. Mol. Genet.* 15 (21), 3195–3205. doi:10.1093/hmg/ddl395

D'Aurizio, R., Pippucci, T., Tattini, L., Giusti, B., Pellegrini, M., and Magi, A. (2016). Enhanced copy number variants detection from whole-exome sequencing data using EXCAVATOR2. *Nucleic Acids Res.* 44 (20), e154. doi:10.1093/nar/gkw695

D'Aurizio, R., Semeraro, R., and Magi, A. (2018). Using XCAVATOR and EXCAVATOR2 to identify CNVs from WGS, WES, and TS data. *Curr. Protoc. Hum. Genet.* 98 (1), e65. doi:10.1002/cphg.65

Delgado-Vega, A. M., Martínez-Bueno, M., Oparina, N. Y., López Herráez, D., Kristjansdottir, H., Steinsson, K., et al. (2018). Whole exome sequencing of patients from multicase families with systemic lupus erythematosus identifies multiple rare variants. *Sci. Rep.* 8 (1), 8775. doi:10.1038/s41598-018-26274-y

Demirkaya, E., Zhou, Q., Smith, C. K., Ombrello, M. J., Deuitch, N., Tsai, W. L., et al. (2017). Brief report: deficiency of complement 1r subcomponent in early-onset systemic

lupus erythematosus: the role of disease-modifying alleles in a monogenic disease. *Arthritis Rheumatol.* 69 (9), 1832–1839. doi:10.1002/art.40158

Deng, Y., and Tsao, B. P. (2014). Advances in lupus genetics and epigenetics. *Curr. Opin. Rheumatol.* 26 (5), 482–492. doi:10.1097/BOR.0000000000000086

Der, E., Suryawanshi, H., Morozov, P., Kustagi, M., Goilav, B., Ranabothu, S., et al. (2019). Tubular cell and keratinocyte single-cell transcriptomics applied to lupus nephritis reveal type I IFN and fibrosis relevant pathways. *Nat. Immunol.* 20 (7), 915–927. doi:10.1038/s41590-019-0386-1

Duncavage, E. J., Schroeder, M. C., O'Laughlin, M., Wilson, R., MacMillan, S., Bohannon, A., et al. (2021). Genome sequencing as an alternative to cytogenetic analysis in myeloid cancers. *N. Engl. J. Med.* 384 (10), 924–935. doi:10.1056/NEJMoa2024534

Elghzaly, A. A., Sun, C., Looger, L. L., Hirose, M., Salama, M., Khalil, N. M., et al. (2022). Genome-wide association study for systemic lupus erythematosus in an egyptian population. *Front. Genet.* 13, 948505. doi:10.3389/fgene.2022.948505

Fang, M., Abolhassani, H., Lim, C. K., Zhang, J., and Hammarstrom, L. (2016). Next generation sequencing data analysis in primary immunodeficiency disorders - future directions. *J. Clin. Immunol.* 36 (Suppl. 1), 68–75. doi:10.1007/s10875-016-0260-y

Fasano, S., Milone, A., Nicoletti, G. F., Isenberg, D. A., and Ciccia, F. (2023). Precision medicine in systemic lupus erythematosus. *Nat. Rev. Rheumatol.* 19 (6), 331–342. doi:10.1038/s41584-023-00948-y

Fernando, M. M., Freudenberg, J., Lee, A., Morris, D. L., Boteva, L., Rhodes, B., et al. (2012). Transancestral mapping of the MHC region in systemic lupus erythematosus identifies new independent and interacting loci at MSH5, HLA-DPB1 and HLA-G. *Ann. Rheum. Dis.* 71 (5), 777–784. doi:10.1136/annrheumdis-2011-200808

Feuk, L., Carson, A. R., and Scherer, S. W. (2006). Structural variation in the human genome. *Nat. Rev. Genet.* 7 (2), 85–97. doi:10.1038/nrg1767

Field, M. A. (2021). Detecting pathogenic variants in autoimmune diseases using high-throughput sequencing. *Immunol. Cell Biol.* 99 (2), 146–156. doi:10.1111/imcb.12372

Fitzgerald, T., and Birney, E. (2022). CNest: a novel copy number association discovery method uncovers 862 new associations from 200,629 whole-exome sequence datasets in the UK Biobank. *Cell Genom* 2 (8), 100167. doi:10.1016/j.xgen.2022.100167

Gabrielaite, M., Torp, M. H., Rasmussen, M. S., Andreu-Sanchez, S., Vieira, F. G., Pedersen, C. B., et al. (2021). A comparison of tools for copy-number variation detection in germline whole exome and whole genome sequencing data. *Cancers (Basel).* 13 (24), 6283. doi:10.3390/cancers13246283

Garcia-Ortiz, H., Velazquez-Cruz, R., Espinosa-Rosales, F., Jimenez-Morales, S., Baca, V., and Orozco, L. (2010). Association of TLR7 copy number variation with susceptibility to childhood-onset systemic lupus erythematosus in Mexican population. *Ann. Rheum. Dis.* 69 (10), 1861–1865. doi:10.1136/ard.2009.124313

Gateva, V., Sandling, J. K., Hom, G., Taylor, K. E., Chung, S. A., Sun, X., et al. (2009). A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. *Nat. Genet.* 41 (11), 1228–1233. doi:10.1038/ng.468

Gaynor, S. M., Joseph, T., Bai, X., Krasheninina, O., Boutkov, B., Maxwell, E., et al. (2023). Yield of genetic association signals from genomes, exomes, and imputations in the UK biobank. medRxiv.

Genomes Project, C., Abecasis, G. R., Altshuler, D., Auton, A., Brooks, L. D., Durbin, R. M., et al. (2010). A map of human genome variation from population-scale sequencing. *Nature* 467 (7319), 1061–1073. doi:10.1038/nature09534

Genomes Project, C., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., et al. (2015). A global reference for human genetic variation. *Nature* 526 (7571), 68–74. doi:10.1038/nature15393

German, D., and Stephan, O. (2019). ClinCNV: novel method for allele-specific somatic copy-number alterations detection. bioRxiv.

Giltiay, N. V., Chappell, C. P., Sun, X., Kolhatkar, N., Teal, T. H., Wiedeman, A. E., et al. (2013). Overexpression of TLR7 promotes cell-intrinsic expansion and autoantibody production by transitional T1 B cells. *J. Exp. Med.* 210 (12), 2773–2789. doi:10.1084/jem.20122798

Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G., et al. (2005). The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 307 (5714), 1434–1440. doi:10.1126/science.1101160

Goodnow, C. C., Sprent, J., Fazekas de St Groth, B., and Vinuesa, C. G. (2005). Cellular and genetic mechanisms of self tolerance and autoimmunity. *Nature* 435 (7042), 590–597. doi:10.1038/nature03724

Goodwin, S., McPherson, J. D., and McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 17 (6), 333–351. doi:10.1038/nrg.2016.49

Graham, R. R., Cotsapas, C., Davies, L., Hackett, R., Lessard, C. J., Leon, J. M., et al. (2008). Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. *Nat. Genet.* 40 (9), 1059–1061. doi:10.1038/ng.200

Gulati, G., and Brunner, H. I. (2018). Environmental triggers in systemic lupus erythematosus. *Semin. Arthritis Rheum.* 47 (5), 710–717. doi:10.1016/j.semarthrit.2017.10.001

Gulilat, M., Lamb, T., Teft, W. A., Wang, J., Dron, J. S., Robinson, J. F., et al. (2019). Targeted next generation sequencing as a tool for precision medicine. *BMC Med. Genomics* 12 (1), 81. doi:10.1186/s12920-019-0527-2

Han, J. W., Zheng, H. F., Cui, Y., Sun, L. D., Ye, D. Q., Hu, Z., et al. (2009). Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat. Genet.* 41 (11), 1234–1237. doi:10.1038/ng.472

Hanaei, S., Sanati, G., Zoghi, S., Gharibzadeh, S., Ziaee, V., and Rezaei, N. (2020). The status of FOXP3 gene methylation in pediatric systemic lupus erythematosus. *Allergol. Immunopathol. Madr.* 48 (4), 332–338. doi:10.1016/j.aller.2020.03.014

Hanaya, A., Miyamae, T., Kishi, T., Sahara, M., Tani, Y., Yamanaka, H., et al. (2017). Systemic lupus erythematosus associated with RASopathy. *Mod. Rheumatol. Case Rep.* 1 (2), 94–98. doi:10.1080/24725625.2017.1337310

Hanscombe, K. B., Morris, D. L., Noble, J. A., Dilthey, A. T., Tombleson, P., Kaufman, K. M., et al. (2018). Genetic fine mapping of systemic lupus erythematosus MHC associations in Europeans and African Americans. *Hum. Mol. Genet.* 27 (21), 3813–3824. doi:10.1093/hmg/ddy280

Harley, I. T. W., and Sawalha, A. H. (2022). Systemic lupus erythematosus as a genetic disease. *Clin. Immunol.* 236, 108953. doi:10.1016/j.clim.2022.108953

Hassan, S., Bahar, R., Johan, M. F., Mohamed Hashim, E. K., Abdullah, W. Z., Esa, E., et al. (2023). Next-generation sequencing (NGS) and third-generation sequencing (TGS) for the diagnosis of thalassemia. *Diagn. (Basel)* 13 (3), 373. doi:10.3390/diagnostics13030373

Hedrich, C. M. (2017). Epigenetics in SLE. *Curr. Rheumatol. Rep.* 19 (9), 58. doi:10.1007/s11926-017-0685-1

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 (9), 900–909. doi:10.1056/NEJMoa0707865

Hong, S. M., Chen, W., Feng, J., Dai, D., and Shen, N. (2022). Novel mutations in ACP5 and SAMHD1 in a patient with pediatric systemic lupus erythematosus. *Front. Pediatr.* 10, 885006. doi:10.3389/fped.2022.885006

Hu, H., Geng, Z., Zhang, S., Xu, Y., Wang, Q., Chen, S., et al. (2023). Rare copy number variation analysis identifies disease-related variants in atrioventricular septal defect patients. *Front. Genet.* 14, 1075349. doi:10.3389/fgene.2023.1075349

Hu, N., Qiu, X., Luo, Y., Yuan, J., Li, Y., Lei, W., et al. (2008). Abnormal histone modification patterns in lupus CD4+ T cells. *J. Rheumatol.* 35 (5), 804–810.

Huang, X., Luu, L. D. W., Jia, N., Zhu, J., Fu, J., Xiao, F., et al. (2022). Multi-platform omics analysis reveals molecular signatures for pathogenesis and activity of systemic lupus erythematosus. *Front. Immunol.* 13, 833699. doi:10.3389/fimmu.2022.833699

Hudson, L. L., Rocca, K., Song, Y. W., and Pandey, J. P. (2002). CTLA-4 gene polymorphisms in systemic lupus erythematosus: a highly significant association with a determinant in the promoter region. *Hum. Genet.* 111 (4-5), 452–455. doi:10.1007/s00439-002-0807-2

Hughes, T., Adler, A., Kelly, J. A., Kaufman, K. M., Williams, A. H., Langefeld, C. D., et al. (2012). Evidence for gene-gene epistatic interactions among susceptibility loci for systemic lupus erythematosus. *Arthritis Rheum.* 64 (2), 485–492. doi:10.1002/art.33354

Human Genome Structural Variation Working, G., Eichler, E. E., Nickerson, D. A., Altshuler, D., Bowcock, A. M., Brooks, L. D., et al. (2007). Completing the map of human genetic variation. *Nature* 447 (7141), 161–165. doi:10.1038/447161a

Iafrate, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y., et al. (2004). Detection of large-scale variation in the human genome. *Nat. Genet.* 36 (9), 949–951. doi:10.1038/ng1416

Jog, N. R., and James, J. A. (2020). Epstein barr virus and autoimmune responses in systemic lupus erythematosus. *Front. Immunol.* 11, 623944. doi:10.3389/fimmu.2020.623944

Jones, S. A., Cantsilieris, S., Fan, H., Cheng, Q., Russ, B. E., Tucker, E. J., et al. (2019). Rare variants in non-coding regulatory regions of the genome that affect gene expression in systemic lupus erythematosus. *Sci. Rep.* 9 (1), 15433. doi:10.1038/s41598-019-51864-9

Jury, E. C., Flores-Borja, F., Kalsi, H. S., Lazarus, M., Isenberg, D. A., Mauri, C., et al. (2010). Abnormal CTLA-4 function in T cells from patients with systemic lupus erythematosus. *Eur. J. Immunol.* 40 (2), 569–578. doi:10.1002/eji.200939781

Kamitaki, N., Sekar, A., Handsaker, R. E., de Rivera, H., Tooley, K., Morris, D. L., et al. (2020). Complement genes contribute sex-biased vulnerability in diverse disorders. *Nature* 582 (7813), 577–581. doi:10.1038/s41586-020-2277-x

Kanai, M., Akiyama, M., Takahashi, A., Matoba, N., Momozawa, Y., Ikeda, M., et al. (2018). Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* 50 (3), 390–400. doi:10.1038/s41588-018-0047-6

Kang, J. H., Xu, H., Choi, S. E., Park, D. J., Lee, J. K., Kwok, S. K., et al. (2020). Obesity increases the incidence of new-onset lupus nephritis and organ damage during follow-up in patients with systemic lupus erythematosus. *Lupus* 29 (6), 578–586. doi:10.1177/0961203320913616

Keshavarz-Fathi, M., Sanati, G., Sadr, M., Mohebbi, B., Ziaee, V., and Rezaei, N. (2022). DNA methylation of CD70 promoter in juvenile systemic lupus erythematosus. *Fetal Pediatr. Pathol.* 41 (1), 58–67. doi:10.1080/15513815.2020.1764681

Khunsriraksakul, C., Li, Q., Markus, H., Patrick, M. T., Sauteraud, R., McGuire, D., et al. (2023). Multi-ancestry and multi-trait genome-wide association meta-analyses inform clinical risk prediction for systemic lupus erythematosus. *Nat. Commun.* 14 (1), 668. doi:10.1038/s41467-023-36306-5

Khunsriraksakul, C., Markus, H., Olsen, N. J., Carrel, L., Jiang, B., and Liu, D. J. (2022). Construction and application of polygenic risk scores in autoimmune diseases. *Front. Immunol.* 13, 889296. doi:10.3389/fimmu.2022.889296

Kilburn, K. H., and Warshaw, R. H. (1992). Prevalence of symptoms of systemic lupus erythematosus (SLE) and of fluorescent antinuclear antibodies associated with chronic exposure to trichloroethylene and other chemicals in well water. *Environ. Res.* 57 (1), 1–9. doi:10.1016/s0013-9351(05)80014-3

Kim, J. H., Jung, S. H., Bae, J. S., Lee, H. S., Yim, S. H., Park, S. Y., et al. (2013). Deletion variants of RABGAP1L, 10q21.3, and C4 are associated with the risk of systemic lupus erythematosus in Korean women. *Arthritis Rheum.* 65 (4), 1055–1063. doi:10.1002/art.37854

Klambauer, G., Schwarzbauer, K., Mayr, A., Clevert, D. A., Mitterecker, A., Bodenhofer, U., et al. (2012). cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic Acids Res.* 40 (9), e69. doi:10.1093/nar/gks003

Kozyrev, S. V., Abelson, A. K., Wojcik, J., Zaghlool, A., Linga Reddy, M. V., Sanchez, E., et al. (2008). Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. *Nat. Genet.* 40 (2), 211–216. doi:10.1038/ng.79

Krumm, N., Sudmant, P. H., Ko, A., O'Roak, B. J., Malig, M., Coe, B. P., et al. (2012). Copy number variation detection and genotyping from exome sequence data. *Genome Res.* 22 (8), 1525–1532. doi:10.1101/gr.138115.112

Kuhlenbaumer, G., Hullmann, J., and Appenzeller, S. (2011). Novel genomic techniques open new avenues in the analysis of monogenic disorders. *Hum. Mutat.* 32 (2), 144–151. doi:10.1002/humu.21400

Kusmirek, W., and Nowak, R. (2022). CNVind: an open source cloud-based pipeline for rare CNVs detection in whole exome sequencing data based on the depth of coverage. *BMC Bioinforma.* 23 (1), 85. doi:10.1186/s12859-022-04617-x

Kwon, Y. C., Chun, S., Kim, K., and Mak, A. (2019). Update on the genetics of systemic lupus erythematosus: genome-wide association studies and beyond. *Cells* 8 (10), 1180. doi:10.3390/cells8101180

Lai, W., Feng, X., Yue, M., Cheung, P. W. H., Choi, V. N. T., Song, Y. Q., et al. (2021). Identification of copy number variants in a southern Chinese cohort of patients with congenital scoliosis. *Genes (Basel)* 12 (8), 1213. doi:10.3390/genes12081213

Lee, J. A., Madrid, R. E., Sperle, K., Ritterson, C. M., Hobson, G. M., Garbern, J., et al. (2006). Spastic paraplegia type 2 associated with axonal neuropathy and apparent PLP1 position effect. *Ann. Neurol.* 59 (2), 398–403. doi:10.1002/ana.20732

Lee, J. Y., Kannan, B., Lim, B. Y., Li, Z., Lim, A. H., Loh, J. W., et al. (2022). The multi-dimensional biomarker landscape in cancer immunotherapy. *Int. J. Mol. Sci.* 23 (14), 7839. doi:10.3390/ijms23147839

Lee, Y. H., Harley, J. B., and Nath, S. K. (2005). CTLA-4 polymorphisms and systemic lupus erythematosus (SLE): a meta-analysis. *Hum. Genet.* 116 (5), 361–367. doi:10.1007/s00439-004-1244-1

Lelieveld, S. H., Spielmann, M., Mundlos, S., Veltman, J. A., and Gilissen, C. (2015). Comparison of exome and genome sequencing technologies for the complete capture of protein-coding regions. *Hum. Mutat.* 36 (8), 815–822. doi:10.1002/humu.22813

Leventopoulos, G., Denayer, E., Makrythanasis, P., Papapolychroniou, C., and Fryssira, H. (2010). Noonan syndrome and systemic lupus erythematosus in a patient with a novel KRAS mutation. *Clin. Exp. Rheumatol.* 28 (4), 556–557.

Levy, S. E., and Myers, R. M. (2016). Advancements in next-generation sequencing. *Annu. Rev. Genomics Hum. Genet.* 17, 95–115. doi:10.1146/annurev-genom-083115-022413

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 (Suppl. l_1), i67–i77. doi:10.1093/rheumatology/kew399

Li, D., Gao, H., Zheng, W., Jin, C., Huang, Y., and Pan, S. (2022). Case report: fetal cervical immature teratoma and copy number variations. *Front. Oncol.* 12, 843268. doi:10.3389/fonc.2022.843268

Li, G., Li, Y., Liu, H., Shi, Y., Guan, W., Zhang, T., et al. (2020). Genetic heterogeneity of pediatric systemic lupus erythematosus with lymphoproliferation. *Med. Baltim.* 99 (20), e20232. doi:10.1097/MD.0000000000020232

Liaw, D., Marsh, D. J., Li, J., Dahia, P. L., Wang, S. I., Zheng, Z., et al. (1997). Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. *Nat. Genet.* 16 (1), 64–67. doi:10.1038/ng0597-64

Lisbona, M. P., Moreno, M., Orellana, C., Gratacos, J., and Larrosa, M. (2009). Noonan syndrome associated with systemic lupus erythematosus. *Lupus* 18 (3), 267–269. doi:10.1177/0961203308094996

Liu, J., and Zhang, H. X. (2013). CTLA-4 polymorphisms and systemic lupus erythematosus: a comprehensive meta-analysis. *Genet. Test. Mol. Biomarkers* 17 (3), 226–231. doi:10.1089/gtmb.2012.0302

Liu, Y., Lei, C., Wang, R., Yang, D., Yang, B., Xu, Y., et al. (2022). Case report: whole-exome sequencing-based copy number variation analysis identified a novel DRC1 homozygous exon deletion in a patient with primary ciliary dyskinesia. *Front. Genet.* 13, 940292. doi:10.3389/fgene.2022.940292

Lood, C., Gullstrand, B., Truedsson, L., Olin, A. I., Alm, G. V., Ronnblom, L., et al. (2009). C1q inhibits immune complex-induced interferon-alpha production in plasmacytoid dendritic cells: a novel link between C1q deficiency and systemic lupus erythematosus pathogenesis. *Arthritis Rheum.* 60 (10), 3081–3090. doi:10.1002/art.24852

Love, M. I., Mysickova, A., Sun, R., Kalscheuer, V., Vingron, M., and Haas, S. A. (2011). Modeling read counts for CNV detection in exome sequencing data. *Stat. Appl. Genet. Mol. Biol.* 10 (1), 52. doi:10.2202/1544-6115.1732

Lovewell, T. R., McDonagh, A. J., Messenger, A. G., Azzouz, M., and Tazi-Ahnini, R. (2015). The AIRE -230Y polymorphism affects AIRE transcriptional activity: potential influence on AIRE function in the thymus. *PLoS One* 10 (5), e0127476. doi:10.1371/journal.pone.0127476

Lund, J. M., Alexopoulou, L., Sato, A., Karow, M., Adams, N. C., Gale, N. W., et al. (2004). Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proc. Natl. Acad. Sci. U. S. A.* 101 (15), 5598–5603. doi:10.1073/pnas.0400937101

Lundtoft, C., Pucholt, P., Martin, M., Bianchi, M., Lundstrom, E., Eloranta, M. L., et al. (2022). Complement C4 copy number variation is linked to SSA/ro and SSB/La autoantibodies in systemic inflammatory autoimmune diseases. *Arthritis Rheumatol.* 74 (8), 1440–1450. doi:10.1002/art.42122

Majewski, J., Schwartzentruber, J., Lalonde, E., Montpetit, A., and Jabado, N. (2011). What can exome sequencing do for you? *J. Med. Genet.* 48 (9), 580–589. doi:10.1136/jmedgenet-2011-100223

Mandelker, D., Schmidt, R. J., Ankala, A., McDonald Gibson, K., Bowser, M., Sharma, H., et al. (2016). Navigating highly homologous genes in a molecular diagnostic setting: a resource for clinical next-generation sequencing. *Genet. Med.* 18 (12), 1282–1289. doi:10.1038/gim.2016.58

Manrai, A. K., Funke, B. H., Rehm, H. L., Olesen, M. S., Maron, B. A., Szolovits, P., et al. (2016). Genetic misdiagnoses and the potential for Health disparities. *N. Engl. J. Med.* 375 (7), 655–665. doi:10.1056/NEJMsa1507092

Marchuk, D. S., Crooks, K., Strande, N., Kaiser-Rogers, K., Milko, L. V., Brandt, A., et al. (2018). Increasing the diagnostic yield of exome sequencing by copy number variant analysis. *PLoS One* 13 (12), e0209185. doi:10.1371/journal.pone.0209185

Martinez-Bueno, M., Oparina, N., Dozmorov, M. G., Marion, M. C., Comeau, M. E., Gilkeson, G., et al. (2018). Trans-Ethnic mapping of BANK1 identifies two independent SLE-risk linkage groups enriched for Co-transcriptional splicing marks. *Int. J. Mol. Sci.* 19 (8), 2331. doi:10.3390/ijms19082331

McCarroll, S. A., Hadnott, T. N., Perry, G. H., Sabeti, P. C., Zody, M. C., Barrett, J. C., et al. (2006). Common deletion polymorphisms in the human genome. *Nat. Genet.* 38 (1), 86–92. doi:10.1038/ng1696

Miller, E. C., and Atkinson, J. P. (2012). Overcoming C2 deficiency. *Clin. Immunol.* 144 (3), 269–271. doi:10.1016/j.clim.2012.07.005

Mina, R., and Brunner, H. I. (2013). Update on differences between childhood-onset and adult-onset systemic lupus erythematosus. *Arthritis Res. Ther.* 15 (4), 218. doi:10.1186/ar4256

Minegishi, N., Nishijima, I., Nobukuni, T., Kudo, H., Ishida, N., Terakawa, T., et al. (2019). Biobank establishment and sample management in the Tohoku medical megabank project. *Tohoku J. Exp. Med.* 248 (1), 45–55. doi:10.1620/tjem.248.45

Montufar-Robles, I., Robles-Garnica, J. C., Cadena-Sandoval, D., Barbosa-Cobos, R. E., Gonzalez-Castillo, D. D., Romero-Diaz, J., et al. (2019). The AIRE Ser196Ser synonymous variant is a risk factor for systemic lupus erythematosus. *Cell Immunol.* 346, 103986. doi:10.1016/j.cellimm.2019.103986

Mueller, M., Barros, P., Witherden, A. S., Roberts, A. L., Zhang, Z., Schaschl, H., et al. (2013). Genomic pathology of SLE-associated copy-number variation at the FCGR2C/FCGR3B/FCGR2B locus. *Am. J. Hum. Genet.* 92 (1), 28–40. doi:10.1016/j.ajhg.2012.11.013

Mulder, N., Abimiku, A., Adebamowo, S. N., de Vries, J., Matimba, A., Olowoyo, P., et al. (2018). H3Africa: current perspectives. *Pharmgenomics Pers. Med.* 11, 59–66. doi:10.2147/PGPM.S141546

Musone, S. L., Taylor, K. E., Lu, T. T., Nititham, J., Ferreira, R. C., Ortmann, W., et al. (2008). Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. *Nat. Genet.* 40 (9), 1062–1064. doi:10.1038/ng.202

Neuman, H., and Koren, O. (2017). The gut microbiota: a possible factor influencing systemic lupus erythematosus. *Curr. Opin. Rheumatol.* 29 (4), 374–377. doi:10.1097/BOR.0000000000000395

Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., et al. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461 (7261), 272–276. doi:10.1038/nature08250

Niederer, H. A., Clatworthy, M. R., Willcocks, L. C., and Smith, K. G. (2010). FcgammaRIIB, FcgammaRIIIB, and systemic lupus erythematosus. *Ann. N. Y. Acad. Sci.* 1183, 69–88. doi:10.1111/j.1749-6632.2009.05132.x

Nosrat zehi, S., Nosrat zehi, M., Atighi, S., Dianat, T., and Kord Tamandani, K. (2021). Promoter methylation and expression status of cytotoxic T-lymphocyte-associated antigen-4 gene in patients with lupus. *J. Epigenetics* 2 (1), 31–38. doi:10.22111/jep.2020.30124.1019

Oftedal, B. E., Berger, A. H., Bruserud, O., Goldfarb, Y., Sulen, A., Breivik, L., et al. (2023). A partial form of AIRE deficiency underlies a mild form of autoimmune polyendocrine syndrome type 1. *J. Clin. Invest.* 133 (21), e169704. doi:10.1172/JCI169704

Olfe, L., von Hardenberg, S., Hofmann, W., Auber, B., Baumann, U., Beier, R., et al. (2023). CTLA-4 insufficiency due to a novel CTLA-4 deletion, identified through copy number variation analysis. *Int. Arch. Allergy Immunol.* 184 (1), 76–84. doi:10.1159/000527051

Papp, O., Doma, V., Gil, J., Marko-Varga, G., Karpati, S., Timar, J., et al. (2021). Organ specific copy number variations in visceral metastases of human melanoma. *Cancers (Basel)* 13 (23), 5984. doi:10.3390/cancers13235984

Parks, C. G., de Souza Espindola Santos, A., Barbhaiya, M., and Costenbader, K. H. (2017). Understanding the role of environmental factors in the development of systemic lupus erythematosus. *Best. Pract. Res. Clin. Rheumatol.* 31 (3), 306–320. doi:10.1016/j.berh.2017.09.005

Pereira, K. M. C., Perazzio, S., Faria, A. G. A., Moreira, E. S., Santos, V. C., Grecco, M., et al. (2019). Impact of C4, C4A and C4B gene copy number variation in the susceptibility, phenotype and progression of systemic lupus erythematosus. *Adv. Rheumatol.* 59 (1), 36. doi:10.1186/s42358-019-0076-6

Petersen, B. S., Fredrich, B., Hoeppner, M. P., Ellinghaus, D., and Franke, A. (2017). Opportunities and challenges of whole-genome and -exome sequencing. *BMC Genet.* 18 (1), 14. doi:10.1186/s12863-017-0479-5

Poole, B. D., Scofield, R. H., Harley, J. B., and James, J. A. (2006). Epstein-Barr virus and molecular mimicry in systemic lupus erythematosus. *Autoimmunity* 39 (1), 63–70. doi:10.1080/08916930500484849

Pullabhatla, V., Roberts, A. L., Lewis, M. J., Mauro, D., Morris, D. L., Odhams, C. A., et al. (2018). *De novo* mutations implicate novel genes in systemic lupus erythematosus. *Hum. Mol. Genet.* 27 (3), 421–429. doi:10.1093/hmg/ddx407

Pullmann, R., Jr., Lukac, J., Skerenova, M., Rovensky, J., Hybenova, J., Melus, V., et al. (1999). Cytotoxic T lymphocyte antigen 4 (CTLA-4) dimorphism in patients with systemic lupus erythematosus. *Clin. Exp. Rheumatol.* 17 (6), 725–729.

Raj, P., Song, R., Zhu, H., Riediger, L., Jun, D. J., Liang, C., et al. (2020). Deep sequencing reveals a DAP1 regulatory haplotype that potentiates autoimmunity in systemic lupus erythematosus. *Genome Biol.* 21 (1), 281. doi:10.1186/s13059-020-02184-z

Rapti, M., Zouaghi, Y., Meylan, J., Ranza, E., Antonarakis, S. E., and Santoni, F. A. (2022). CoverageMaster: comprehensive CNV detection and visualization from NGS short reads for genetic medicine applications. *Brief. Bioinform* 23 (2), bbac049. doi:10.1093/bib/bbac049

Raupov, R. K., Suspitsin, E. N., Imelbaev, A. I., and Kostik, M. M. (2022). Simultaneous Onset of Pediatric Systemic Lupus Erythematosus in Twin Brothers: Case Report. *Front. Pediatr.* 10, 929358. doi:10.3389/fped.2022.929358

Ravenscroft, J. C., Suri, M., Rice, G. I., Szynkiewicz, M., and Crow, Y. J. (2011). Autosomal dominant inheritance of a heterozygous mutation in SAMHD1 causing familial chilblain lupus. *Am. J. Med. Genet. A* 155A (1), 235–237. doi:10.1002/ajmg.a.33778

Refeat, M. M., Naggar, W. E., Saied, M. M. E., and Kilany, A. (2023). Whole exome screening of neurodevelopmental regression disorders in a cohort of Egyptian patients. *Neurogenetics* 24 (1), 17–28. doi:10.1007/s10048-022-00703-7

Rehm, H. L., Berg, J. S., Brooks, L. D., Bustamante, C. D., Evans, J. P., Landrum, M. J., et al. (2015). ClinGen--the clinical genome resource. *N. Engl. J. Med.* 372 (23), 2235–2242. doi:10.1056/NEJMsr1406261

Reid, S., Alexsson, A., Frodlund, M., Morris, D., Sandling, J. K., Bolin, K., et al. (2020). High genetic risk score is associated with early disease onset, damage accrual and decreased survival in systemic lupus erythematosus. *Ann. Rheum. Dis.* 79 (3), 363–369. doi:10.1136/annrheumdis-2019-216227

Remmers, E. F., Plenge, R. M., Lee, A. T., Graham, R. R., Hom, G., Behrens, T. W., et al. (2007). STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N. Engl. J. Med.* 357 (10), 977–986. doi:10.1056/NEJMoa073003

Reynier, F., Petit, F., Paye, M., Turrel-Davin, F., Imbert, P. E., Hot, A., et al. (2011). Importance of correlation between gene expression levels: application to the type I interferon signature in rheumatoid arthritis. *PLoS One* 6 (10), e24828. doi:10.1371/journal.pone.0024828

Rice, G. I., Del Toro Duany, Y., Jenkinson, E. M., Forte, G. M., Anderson, B. H., Ariaudo, G., et al. (2014). Gain-of-function mutations in IFIH1 cause a spectrum of human disease phenotypes associated with upregulated type I interferon signaling. *Nat. Genet.* 46 (5), 503–509. doi:10.1038/ng.2933

Rice, G. I., Kasher, P. R., Forte, G. M., Mannion, N. M., Greenwood, S. M., Szynkiewicz, M., et al. (2012). Mutations in ADAR1 cause Aicardi-Goutieres syndrome associated with a type I interferon signature. *Nat. Genet.* 44 (11), 1243–1248. doi:10.1038/ng.2414

Rice, G. I., Rodero, M. P., and Crow, Y. J. (2015). Human disease phenotypes associated with mutations in TREX1. *J. Clin. Immunol.* 35 (3), 235–243. doi:10.1007/s10875-015-0147-3

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

Rioux, J. D., Goyette, P., Vyse, T. J., Hammarström, L., Fernando, M. M., Green, T., et al. (2009). Mapping of multiple susceptibility variants within the MHC region for 7 immune-mediated diseases. *Proc. Natl. Acad. Sci. U S A* 106 (44), 18680–18685. doi:10.1073/pnas.0909307106

Royer-Bertrand, B., Cisarova, K., Niel-Butschi, F., Mittaz-Crettol, L., Fodstad, H., and Superti-Furga, A. (2021). CNV detection from exome sequencing data in routine diagnostics of rare genetic disorders: opportunities and limitations. *Genes (Basel).* 12 (9), 1427. doi:10.3390/genes12091427

Sandling, J. K., Pucholt, P., Hultin Rosenberg, L., Farias, F. H. G., Kozyrev, S. V., Eloranta, M. L., et al. (2021). Molecular pathways in patients with systemic lupus erythematosus revealed by gene-centred DNA sequencing. *Ann. Rheum. Dis.* 80 (1), 109–117. doi:10.1136/annrheumdis-2020-218636

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 74 (12), 5463–5467. doi:10.1073/pnas.74.12.5463

Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., et al. (2023). Next-generation sequencing technology: current trends and advancements. *Biol. (Basel)* 12 (7), 997. doi:10.3390/biology12070997

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., et al. (2004). Large-scale copy number polymorphism in the human genome. *Science* 305 (5683), 525–528. doi:10.1126/science.1098918

Sestak, A. L., Furnrohr, B. G., Harley, J. B., Merrill, J. T., and Namjou, B. (2011). The genetics of systemic lupus erythematosus and implications for targeted therapy. *Ann. Rheum. Dis.* 70 (Suppl. 1), i37–i43. doi:10.1136/ard.2010.138057

Shang, S., Zhou, Y., Chen, K., Chen, L., Li, P., Li, D., et al. (2022). A novel gene CDC27 causes SLE and is associated with the disease activity. *Front. Immunol.* 13, 876963. doi:10.3389/fimmu.2022.876963

Shaw, T., Fok, R., Courtney, E., Li, S. T., Chiang, J., and Ngeow, J. (2023). Missed diagnosis or misdiagnosis: common pitfalls in genetic testing. *Singap. Med. J.* 64 (1), 67–73. doi:10.4103/singaporemedj.SMJ-2021-467

Shibata, T., Ohto, U., Nomura, S., Kibata, K., Motoi, Y., Zhang, Y., et al. (2016). Guanosine and its modified derivatives are endogenous ligands for TLR7. *Int. Immunol.* 28 (5), 211–222. doi:10.1093/intimm/dxv062

Slatko, B. E., Gardner, A. F., and Ausubel, F. M. (2018). Overview of next-generation sequencing technologies. *Curr. Protoc. Mol. Biol.* 122 (1), e59. doi:10.1002/cpmb.59

Sohail, M., Palma-Martinez, M. J., Chong, A. Y., Quinto-Cortes, C. D., Barberena-Jonas, C., Medina-Munoz, S. G., et al. (2023). Mexican Biobank advances population and medical genomics of diverse ancestries. *Nature* 622, 775–783. doi:10.1038/s41586-023-06560-0

Somerville, M. J., Mervis, C. B., Young, E. J., Seo, E. J., del Campo, M., Bamforth, S., et al. (2005). Severe expressive-language delay related to duplication of the Williams-Beuren locus. *N. Engl. J. Med.* 353 (16), 1694–1701. doi:10.1056/NEJMoa051962

Speyer, C. B., and Costenbader, K. H. (2018). Cigarette smoking and the pathogenesis of systemic lupus erythematosus. *Expert Rev. Clin. Immunol.* 14 (6), 481–487. doi:10.1080/1744666X.2018.1473035

Sudmant, P. H., Rausch, T., Gardner, E. J., Handsaker, R. E., Abyzov, A., Huddleston, J., et al. (2015). An integrated map of structural variation in 2,504 human genomes. *Nature* 526 (7571), 75–81. doi:10.1038/nature15394

Taheri, M., Eghtedarian, R., Dinger, M. E., and Ghafouri-Fard, S. (2020). Exploring the role of non-coding RNAs in the pathophysiology of systemic lupus erythematosus. *Biomolecules* 10 (6), 937. doi:10.3390/biom10060937

Takeda, K., and Akira, S. (2005). Toll-like receptors in innate immunity. *Int. Immunol.* 17 (1), 1–14. doi:10.1093/intimm/dxh186

Talevich, E., Shain, A. H., Botton, T., and Bastian, B. C. (2016). CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput. Biol.* 12 (4), e1004873. doi:10.1371/journal.pcbi.1004873

Tam, V., Patel, N., Turcotte, M., Bosse, Y., Pare, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* 20 (8), 467–484. doi:10.1038/s41576-019-0127-1

Tang, Y., and Luo, Y. (2022). Identification of a novel mutation in complement receptor 2 in Chinese familial systemic lupus erythematosus. *Arch. Rheumatol.* 37 (4), 566–573. doi:10.46497/ArchRheumatol.2022.9167

Terao, C., Yamada, R., Ohmura, K., Takahashi, M., Kawaguchi, T., Kochi, Y., et al. (2011). The human AIRE gene at chromosome 21q22 is a genetic determinant for the predisposition to rheumatoid arthritis in Japanese population. *Hum. Mol. Genet.* 20 (13), 2680–2685. doi:10.1093/hmg/ddr161

Tirosh, I., Spielman, S., Barel, O., Ram, R., Stauber, T., Paret, G., et al. (2019). Whole exome sequencing in childhood-onset lupus frequently detects single gene etiologies. *Pediatr. Rheumatol. Online J.* 17 (1), 52. doi:10.1186/s12969-019-0349-y

Torkamani, A., Wineinger, N. E., and Topol, E. J. (2018). The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* 19 (9), 581–590. doi:10.1038/s41576-018-0018-x

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

Tuzun, E., Sharp, A. J., Bailey, J. A., Kaul, R., Morrison, V. A., Pertz, L. M., et al. (2005). Fine-scale structural variation of the human genome. *Nat. Genet.* 37 (7), 727–732. doi:10.1038/ng1562

Uehara, T., Hosogaya, N., Matsuo, N., and Kosaki, K. (2018). Systemic lupus erythematosus in a patient with Noonan syndrome-like disorder with loose anagen hair 1: more than a chance association. *Am. J. Med. Genet. A* 176 (7), 1662–1666. doi:10.1002/ajmg.a.38834

Usher, C. L., and McCarroll, S. A. (2015). Complex and multi-allelic copy number variation in human disease. *Brief. Funct. Genomics* 14 (5), 329–338. doi:10.1093/bfgp/elv028

Van Coillie, S., Wiernicki, B., and Xu, J. (2020). Molecular and cellular functions of CTLA-4. *Adv. Exp. Med. Biol.* 1248, 7–32. doi:10.1007/978-981-15-3266-5_2

van Dijk, E. L., Auger, H., Jaszczyszyn, Y., and Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends Genet.* 30 (9), 418–426. doi:10.1016/j.tig.2014.07.001

Van Hout, C. V., Tachmazidou, I., Backman, J. D., Hoffman, J. D., Liu, D., Pandey, A. K., et al. (2020). Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* 586 (7831), 749–756. doi:10.1038/s41586-020-2853-0

Villarino, A. V., Kanno, Y., and O'Shea, J. J. (2017). Mechanisms and consequences of Jak-STAT signaling in the immune system. *Nat. Immunol.* 18 (4), 374–384. doi:10.1038/ni.3691

Vinuesa, C. G., Shen, N., and Ware, T. (2023). Genetics of SLE: mechanistic insights from monogenic disease and disease-associated variants. *Nat. Rev. Nephrol.* 19 (9), 558–572. doi:10.1038/s41581-023-00732-x

Vordenbaumen, S., Rosenbaum, A., Gebhard, C., Raithel, J., Sokolowski, A., Dusing, C., et al. (2021). Associations of site-specific CD4^{+}-T-cell hypomethylation within CD40-ligand promotor and enhancer regions with disease activity of women with systemic lupus erythematosus. *Lupus* 30 (1), 45–51. doi:10.1177/0961203320965690

Wahren-Herlenius, M., and Dorner, T. (2013). Immunopathogenic mechanisms of systemic autoimmune disease. *Lancet* 382 (9894), 819–831. doi:10.1016/S0140-6736(13)60954-X

Wang, Q., Qin, T., Wang, X., Li, J., Lin, X., Wang, D., et al. (2022). Whole-exome sequencing and copy number analysis in a patient with warburg micro syndrome. *Genes (Basel)* 13 (12), 2364. doi:10.3390/genes13122364

Wang, Y. F., Zhang, Y., Lin, Z., Zhang, H., Wang, T. Y., 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

Webb, R., Kelly, J. A., Somers, E. C., Hughes, T., Kaufman, K. M., Sanchez, E., et al. (2011). Early disease onset is predicted by a higher genetic risk for lupus and is associated with a more severe phenotype in lupus patients. *Ann. Rheum. Dis.* 70 (1), 151–156. doi:10.1136/ard.2010.141697

Wei, W. H., Hemani, G., and Haley, C. S. (2014). Detecting epistasis in human complex traits. *Nat. Rev. Genet.* 15 (11), 722–733. doi:10.1038/nrg3747

Willcocks, L. C., Lyons, P. A., Clatworthy, M. R., Robinson, J. I., Yang, W., Newland, S. A., et al. (2008). Copy number of FCGR3B, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake. *J. Exp. Med.* 205 (7), 1573–1582. doi:10.1084/jem.20072413

Wu, C. W., Lim, T. Y., Wang, C., Seltzsam, S., Zheng, B., Schierbaum, L., et al. (2022). Copy number variation analysis facilitates identification of genetic causation in patients with congenital anomalies of the kidney and urinary tract. *Eur. Urol. Open Sci.* 44, 106–112. doi:10.1016/j.euros.2022.08.004

Xiao, T., and Zhou, W. (2020). The third generation sequencing: the advanced approach to genetic diseases. *Transl. Pediatr.* 9 (2), 163–173. doi:10.21037/tp.2020.03.06

Yang, S., Fujikado, N., Kolodin, D., Benoist, C., and Mathis, D. (2015). Immune tolerance. Regulatory T cells generated early in life play a distinct role in maintaining self-tolerance. *Science* 348 (6234), 589–594. doi:10.1126/science.aaa7017

Yang, W., Shen, N., Ye, D. Q., Liu, Q., Zhang, Y., Qian, X. X., et al. (2010). Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. *PLoS Genet.* 6 (2), e1000841. doi:10.1371/journal.pgen.1000841

Yang, Y., Chung, E. K., Wu, Y. L., Savelli, S. L., Nagaraja, H. N., Zhou, B., et al. (2007). Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am. J. Hum. Genet.* 80 (6), 1037–1054. doi:10.1086/518257

Yasutomo, K., Horiuchi, T., Kagami, S., Tsukamoto, H., Hashimura, C., Urushihara, M., et al. (2001). Mutation of DNASE1 in people with systemic lupus erythematosus. *Nat. Genet.* 28 (4), 313–314. doi:10.1038/91070

Yaung, K., Yeo, J., Kumar, P., Wasser, M., Chew, M., Ravelli, A., et al. (2023). Artificial intelligence and high-dimensional technologies in the theragnosis of systemic lupus erythematosus. *Lancet Rheumatology* 5, e151–e165. doi:10.1016/S2665-9913(23)00010-3

You, Q., Yang, X., Peng, Z., Xu, L., and Wang, J. (2018). Development and applications of a high throughput genotyping tool for polyploid crops: single nucleotide polymorphism (SNP) array. *Front. Plant Sci.* 9, 104. doi:10.3389/fpls.2018.00104

Zawistowski, M., Fritsche, L. G., Pandit, A., Vanderwerff, B., Patil, S., Schmidt, E. M., et al. (2023). The Michigan Genomics Initiative: a biobank linking genotypes and electronic clinical records in Michigan Medicine patients. *Cell Genom* 3 (2), 100257. doi:10.1016/j.xgen.2023.100257

Zhang, M., Gu, Y., Huang, S., Lou, Q., Xie, Q., Xu, Z., et al. (2019). Copy number variations and polymorphisms in HSP90AB1 and risk of systemic lupus erythematosus and efficacy of glucocorticoids. *J. Cell Mol. Med.* 23 (8), 5340–5348. doi:10.1111/jcmm.14410

Zhang, Q., Qin, Z., Yi, S., Wei, H., Zhou, X. Z., and Su, J. (2021). Clinical application of whole-exome sequencing: a retrospective, single-center study. *Exp. Ther. Med.* 22 (1), 753. doi:10.3892/etm.2021.10185

Zhang, Z., Ohto, U., Shibata, T., Krayukhina, E., Taoka, M., Yamauchi, Y., et al. (2016). Structural analysis reveals that toll-like receptor 7 is a dual receptor for guanosine and single-stranded RNA. *Immunity* 45 (4), 737–748. doi:10.1016/j.immuni.2016.09.011

Zhang, Z., Ohto, U., Shibata, T., Taoka, M., Yamauchi, Y., Sato, R., et al. (2018). Structural analyses of toll-like receptor 7 reveal detailed RNA sequence specificity and recognition mechanism of agonistic ligands. *Cell Rep.* 25 (12), 3371–3381. doi:10.1016/j.celrep.2018.11.081

Zhao, L., Liu, H., Yuan, X., Gao, K., and Duan, J. (2020). Comparative study of whole exome sequencing-based copy number variation detection tools. *BMC Bioinforma.* 21 (1), 97. doi:10.1186/s12859-020-3421-1

Zhou, J., Zhang, M., Li, X., Wang, Z., Pan, D., and Shi, Y. (2021). Performance comparison of four types of target enrichment baits for exome DNA sequencing. *Hereditas* 158 (1), 10. doi:10.1186/s41065-021-00171-3

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 06 March 2024
DOI 10.3389/fgene.2024.1296570

OPEN ACCESS

EDITED BY
Che Kang Lim,
Singapore General Hospital, Singapore

REVIEWED BY
Sammed Mandape,
Othram Inc., United States
Chunying Li,
Georgia State University, United States
Joaquim Carreras,
Tokai University, Japan

*CORRESPONDENCE
Junjie Hou,
houjunjie1979@163.com
Zhixin Yuan,
doctorzx2023@163.com

RECEIVED 19 September 2023
ACCEPTED 16 February 2024
PUBLISHED 06 March 2024

CITATION
Ou W, Qi Z, Liu N, Zhang J, Mi X, Song Y, Fang Y, Cui B, Hou J and Yuan Z (2024), Elucidating the role of TWIST1 in ulcerative colitis: a comprehensive bioinformatics and machine learning approach.
Front. Genet. 15:1296570.
doi: 10.3389/fgene.2024.1296570

Elucidating the role of TWIST1 in ulcerative colitis: a comprehensive bioinformatics and machine learning approach

Wenjie Ou[1], Zhaoxue Qi[2], Ning Liu[3], Junzi Zhang[1], Xuguang Mi[4], Yuan Song[5], Yanqiu Fang[4], Baiying Cui[1], Junjie Hou[6]* and Zhixin Yuan[7]*

[1]School of Clinical Medicine, Changchun University of Chinese Medicine, Changchun, Jilin, China, [2]Department of Secretory Metabolism, The First Hospital of Jilin University, Changchun, Jilin, China, [3]General Surgery of The First Clinical Hospital of Jilin Academy of Chinese Medicine Sciences, Changchun, Jilin, China, [4]Department of Central Laboratory, Jilin Provincial People's Hospital, Changchun, Jilin, China, [5]Department of Gastroenterology, Jilin Provincial People's Hospital, Changchun, Jilin, China, [6]Department of Comprehensive Oncology, Jilin Provincial People's Hospital, Changchun, Jilin, China, [7]Department of Emergency Surgery, Jilin Provincial People's Hospital, Changchun, Jilin, China

Background: Ulcerative colitis (UC) is a common and progressive inflammatory bowel disease primarily affecting the colon and rectum. Prolonged inflammation can lead to colitis-associated colorectal cancer (CAC). While the exact cause of UC remains unknown, this study aims to investigate the role of the TWIST1 gene in UC.

Methods: Second-generation sequencing data from adult UC patients were obtained from the Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) were identified, and characteristic genes were selected using machine learning and Lasso regression. The Receiver Operating Characteristic (ROC) curve assessed TWIST1's potential as a diagnostic factor (AUC score). Enriched pathways were analyzed, including Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Set Variation Analysis (GSVA). Functional mechanisms of marker genes were predicted, considering immune cell infiltration and the competing endogenous RNA (ceRNA) network.

Results: We found 530 DEGs, with 341 upregulated and 189 downregulated genes. TWIST1 emerged as one of four potential UC biomarkers via machine learning. TWIST1 expression significantly differed in two datasets, GSE193677 and GSE83687, suggesting its diagnostic potential (AUC = 0.717 in GSE193677, AUC = 0.897 in GSE83687). Enrichment analysis indicated DEGs associated with TWIST1 were involved in processes like leukocyte migration, humoral immune

Abbreviations: AUC, Area under the curve; IBD, Inflammatory bowel diseases; GEO, Gene Expression Omnibus; DEGs, Differentially expressed genes; ceRNA, Competing endogenous RNA; FDR, False discovery rate; GO, Gene ontology analysis; BP, Biological processes; CC, Cellular components; MF, Molecular functions; KEGG, Kyoto Encyclopedia of genes and Genomes; GSVA, Gene Set Variation Analysis; ROC, Receiver Operating Characteristic; lncRNA, Long non-coding RNA; NK, Natural killing; FC Fold Change; CAC, colitis-associated colorectal cancer; SVM-RFE, Support Vector Machine with Recursive Feature Elimination; LASSO, Least Absolute Shrinkage and Selection Operator; S100P, S100 Calcium Binding Protein P; GPR15, G protein-coupled receptor 15; TWIST1, Twist Family BHLH Transcription Factor 1; RND1, Rho Family GTPase 1; bHLH, basic helix-loop-helix; UC, Ulcerative colitis; .

response, and cell chemotaxis. Immune cell infiltration analysis revealed higher rates of M0 macrophages and resting NK cells in the high TWIST1 expression group, while TWIST1 expression correlated positively with M2 macrophages and resting NK cell infiltration. We constructed a ceRNA regulatory network involving 1 mRNA, 7 miRNAs, and 32 long non-coding RNAs (lncRNAs) to explore TWIST1's regulatory mechanism.

Conclusion: TWIST1 plays a significant role in UC and has potential as a diagnostic marker. This study sheds light on UC's molecular mechanisms and underscores TWIST1's importance in its progression. Further research is needed to validate these findings in diverse populations and investigate TWIST1 as a therapeutic target in UC.

KEYWORDS

ulcerative colitis (UC), Twist1, bioinformatics, machine learning, gene expression omnibus (GEO) database, differentially expressed genes (DEGs), diagnostic marker

1 Introduction

Ulcerative Colitis (UC) is a chronic form of Inflammatory Bowel Disease (IBD) primarily affecting the colon and rectum. The exact cause of UC remains elusive, although several factors, including genetics, environmental triggers, and immune responses, are believed to play pivotal roles in its onset (Kobayashi et al., 2020). While the incidence of adult UC in Asia has historically been relatively low, there has been a noticeable increase in recent years (Du and Ha, 2020). UC not only significantly impairs the quality of life for affected individuals but, in severe cases, also raises the risk of developing colitis-associated colorectal cancer (CAC) (Yashiro, 2014). Hence, the exploration of potential risk markers highly correlated with the occurrence and progression of UC is of paramount importance.

TWIST1, a basic helix-loop-helix (bHLH) transcription factor, was initially identified during embryonic development and plays a pivotal role in cellular migration, differentiation, and morphogenesis (Murre et al., 1989; Jan and January 1993; Kadesch, 1993). In oncological research, the Twist1 gene has garnered significant attention due to its cardinal role in tumor invasion and metastasis (Ren et al., 2016; Ghafouri-Fard et al., 2021). However, the implications of the Twist1 gene in UC remain largely uncharted. A study from June 2018 highlighted that the expression of TWIST1 protein was markedly elevated in tissues from both UC and CAC, and it was closely associated with tissue cellular apoptosis (Anonymous, 2023). Limitations of this study include the exclusive use of immunohistochemistry techniques to investigate gene expression levels within tissues. As a result, it did not explore the correlation between TWIST1 expression and immune factors closely associated with the occurrence and development of UC and CAC. Furthermore, it did not investigate the relationship between TWIST1 expression and the activity of UC. The search for transcriptional regulators of TWIST1 and the exploration of its regulatory targets were also omitted, although these aspects are considered indispensable.

With the advent of bioinformatics and high-throughput sequencing technologies, researchers have pinpointed several genes and pathways intrinsically linked to UC, offering fresh insights into its intricate pathophysiological mechanisms (Kakiuchi et al., 2020; Tong et al., 2021; Xu et al., 2022). Bioinformatics provides a robust analytical framework for identifying pivotal genes associated with UC and analyzing their expression significance. This study harnesses the second-generation sequencing data of adult UC from the Gene Expression Omnibus (GEO) database to probe potential aberrations in the expression levels of the TWIST1 gene. Additionally, the burgeoning field of machine learning bestows capabilities in predictive modeling and pattern discernment, proving indispensable in the interpretation of multifaceted biological datasets. Functional enrichment analysis further facilitates a profound comprehension of the biological intricacies of genes. By juxtaposing differentially expressed genes (DEGs) with functional databases, we endeavor to unveil the biological pathways and processes in which TWIST1 partakes, thereby delving into its prospective role in UC pathogenesis. This research aspires to elucidate the nexus between the TWIST1 gene and UC, fortifying our understanding of its disease mechanisms and informing future therapeutic paradigms.

2 Materials and methods

2.1 Selection and download of the UC dataset

We retrieved matrix files from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) that contained samples of normal human intestinal mucosal tissue and intestinal mucosal tissue from adult patients with UC. Our selection process followed specific criteria: (1) The data pertained to high-throughput sequencing expression profiles of *Homo sapiens*; (2) The samples included biopsied intestinal mucosal tissue from both healthy adults and UC patients; (3) Samples were taken from patients with active clinical disease; (4) Each dataset contained over 6 samples; (5) All the included samples had not been subjected to drug treatment; (6) The dataset provided comprehensive information about each sample. As a result, we identified two datasets for our study. The first, GSE193677 (Argmann et al., 2023), encompassed a total of 461 samples from healthy human subjects (control group) and 126 samples from patients with UC (treatment group). Furthermore, for subsequent validation, we opted for the GSE83687 (Peters et al., 2017) datasets, consisting of 60 samples from healthy human colon tissue and 32 samples from UC-affected colon tissue, as depicted in Table 1. It is worth noting that data from the GEO database is readily accessible to the public, obviating the need for local ethics committee approval.

TABLE 1 Information for selected microarray datasets.

GEO accession	Samples		Country	Attribute
	Con	UC		
GSE193677	461	126	United States	Test set
GSE83687	60	32	United States	Validation set

2.2 Correction, screening and visualization of differentially expressed genes

After downloading the matrix files from the GEO database, we proceeded to process and annotate them utilizing both Perl language (version 32), R language (version 4.30), and Excel. DEGs were derived by subjecting the sample data to filtration through the R limma package. Our filtration criteria encompassed |Log Fold Change (FC)| > 1, and the ensuing *p*-values underwent correction while controlling the false discovery rate (FDR), resulting in an adjusted *p*-value (Q value) < 0.05. Subsequently, the chosen DEGs were subject to visualization and analysis, and the outcome was the generation of heat maps and volcano plots.

2.3 Utilizing machine learning for the identification of disease-related genes

We proceeded to employ machine learning techniques for the additional screening of the acquired DEGs, with the objective of pinpointing genes with a high degree of association with UC. TWO distinct machine learning algorithms, namely, the Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani, 1996) and the Support Vector Machine with Recursive Feature Elimination (SVM-RFE) (Suykens and Vandewalle, 1999), were employed to effectively sift through the pool of DEGs. Finally, R venn package was use to obtain their intersection genes. This enabled us to pinpoint potential disease biomarkers with remarkable precision.

2.4 Validation of TWIST1 expression and diagnostic value

In the GSE193677 and GSE83687 datasets, *t*-test was employed to compare the expression levels of TWIST1 between the UC experimental group and the control group. Sensitivity and specificity of TWIST1 were determined through Receiver

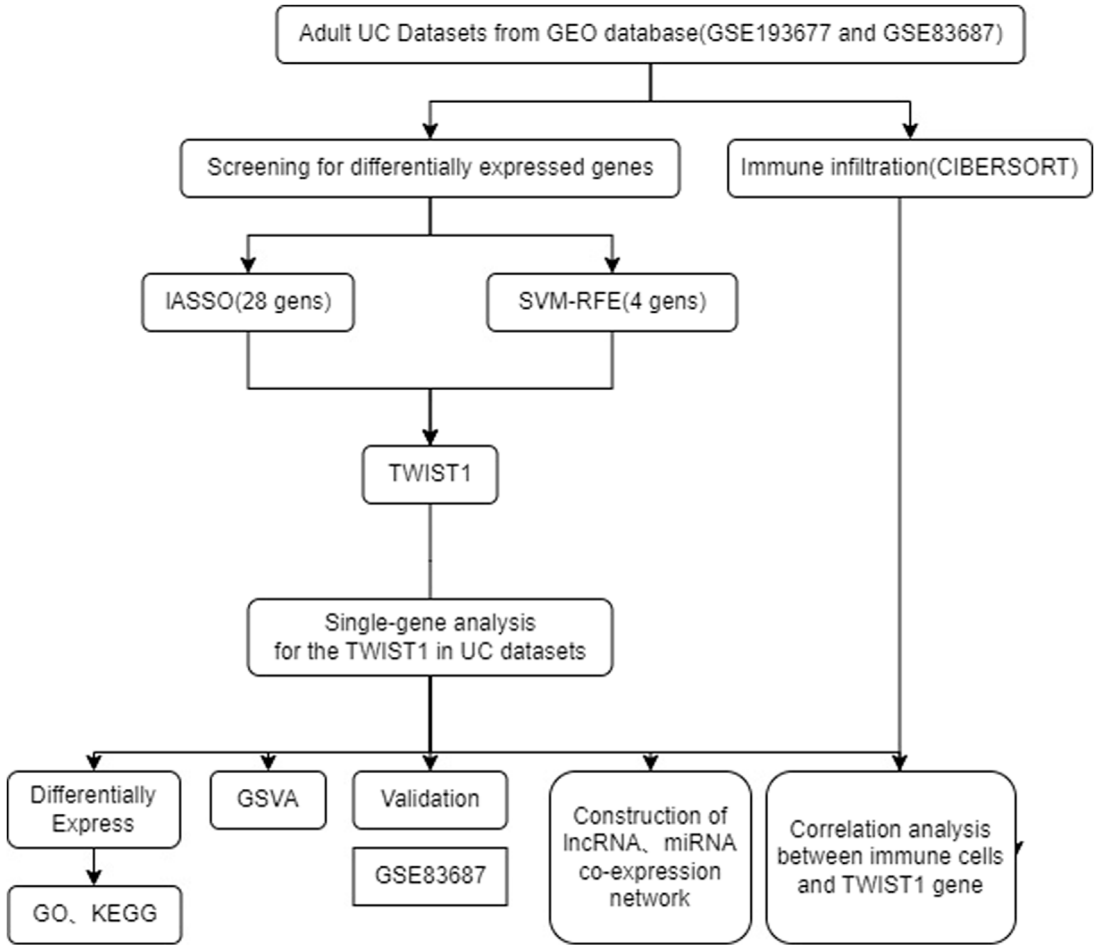


FIGURE 1
The flowchart of the analysis.

10.3389/fgene.2024.1296570

Operating Characteristic (ROC) (Kumar and Indrayan, 2011) curve analysis using the R pROC package. These results were visually depicted using the R ggplot2 package.

Furthermore, Clinical information data for the GSE193677 dataset were obtained, and clinical disease activity was categorized as active or inactive. Kruskal–Wallis tests were

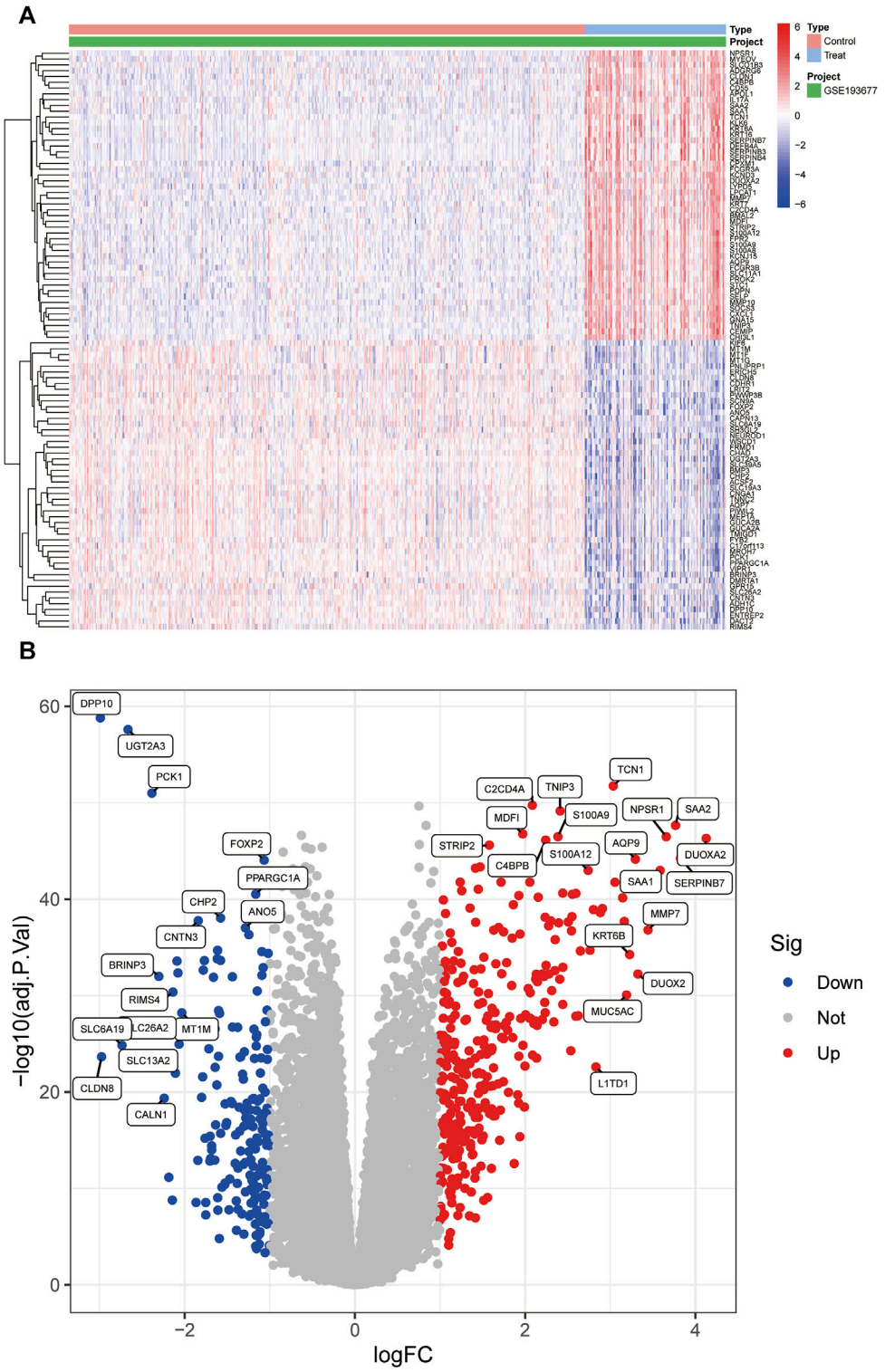


FIGURE 2
The heatmap and a volcano plots. **(A)** The heatmap of DEGs distribution; **(B)** The volcano plots of DEGs. Red represented a high expression of DEG, while blue represented a low expression of DEG.

10.3389/fgene.2024.1296570

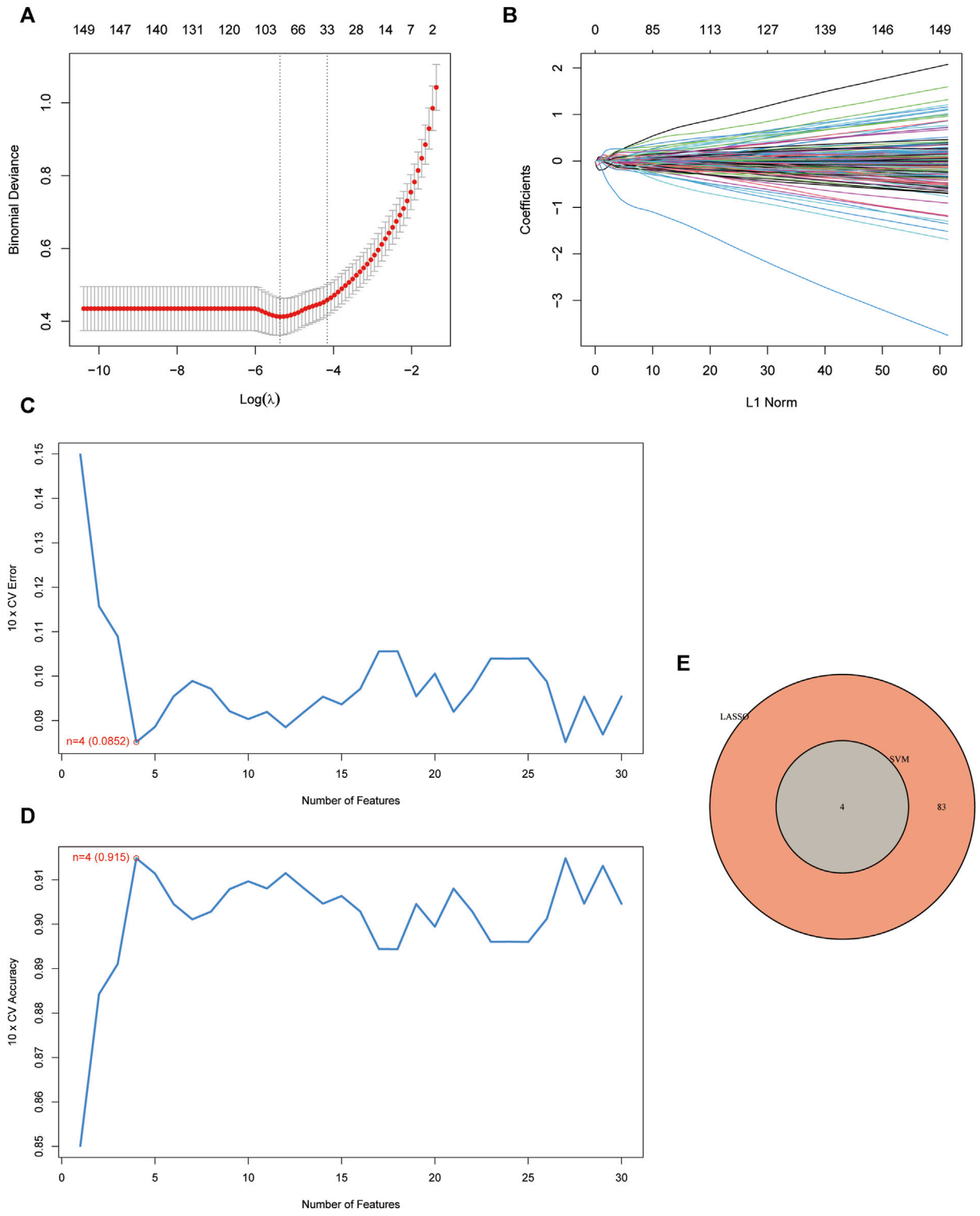


FIGURE 3
Screening of disease-related genes by machine learning. **(A, B)** Regression coefficient path diagram and cross-validation curves in LASSO logistic regression algorithm.; **(C, D)** The curve of change in the predicted true and error value of each gene in SVM-RFE algorithm.; **(E)** Venn diagram demonstrates the intersection of diagnostic markers obtained from the three algorithms.

10.3389/fgene.2024.1296570

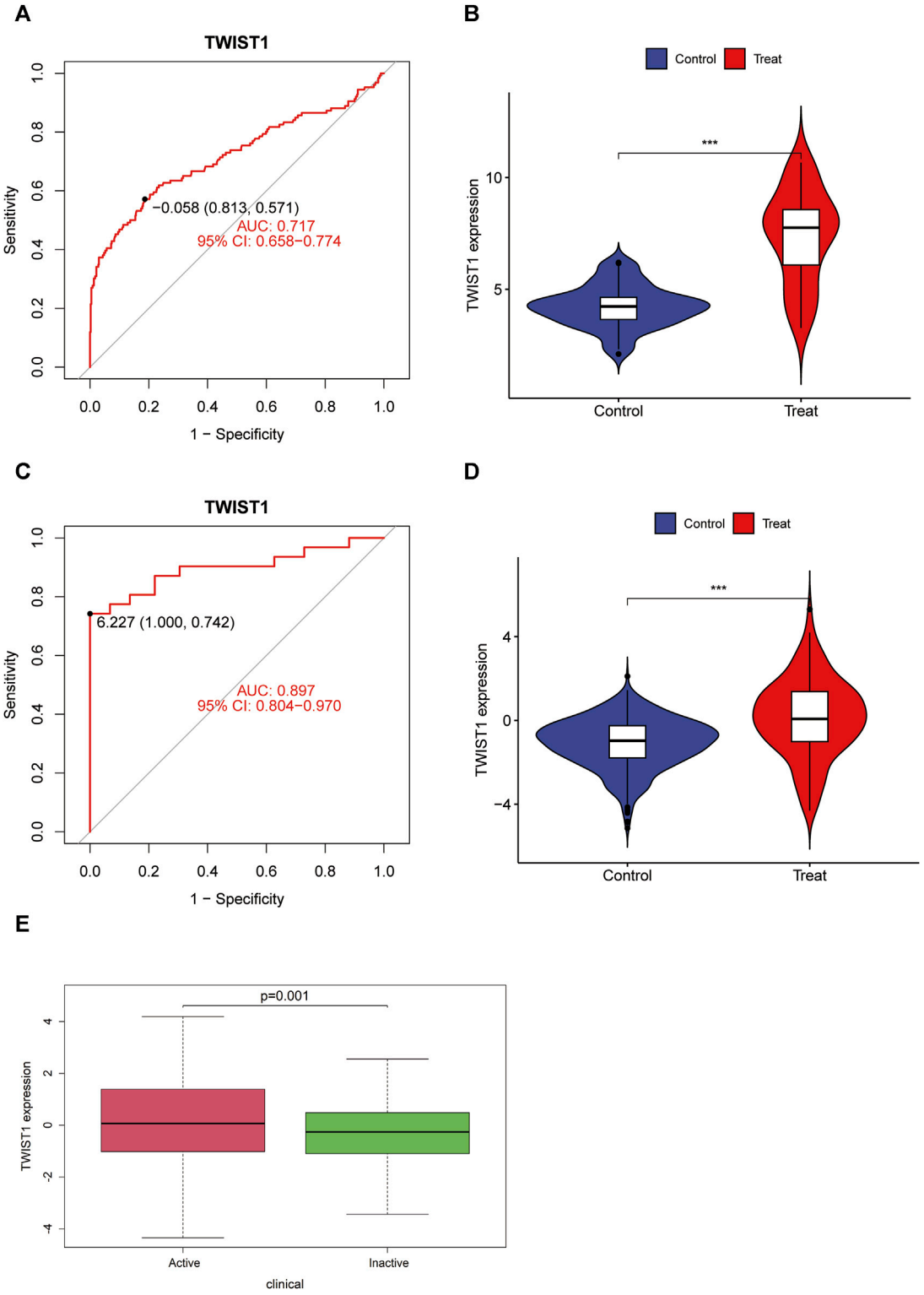


FIGURE 4
Receiver operating characteristic (ROC) curve, expression difference of TWIST1 gene and correlation between TWIST expression and disease activity in UC. **(A, B)** ROC curve and differential expression in GSE193677; **(C, D)** ROC curve and differential expression in GSE83687; **(E)** Boxplot of correlation between TWIST and UC clinical manifestation activity.

conducted to assess the association between clinical disease activity and TWIST1 expression levels in both the UC experimental group and the control group. Statistical significance was defined when the *p*-values from both tests were below 0.05.

2.5 Difference analysis based on the median value of TWIST1 gene expression

Within GSE193677, division into two distinct groups was undertaken based on the distinct levels of TWIST1 expression.

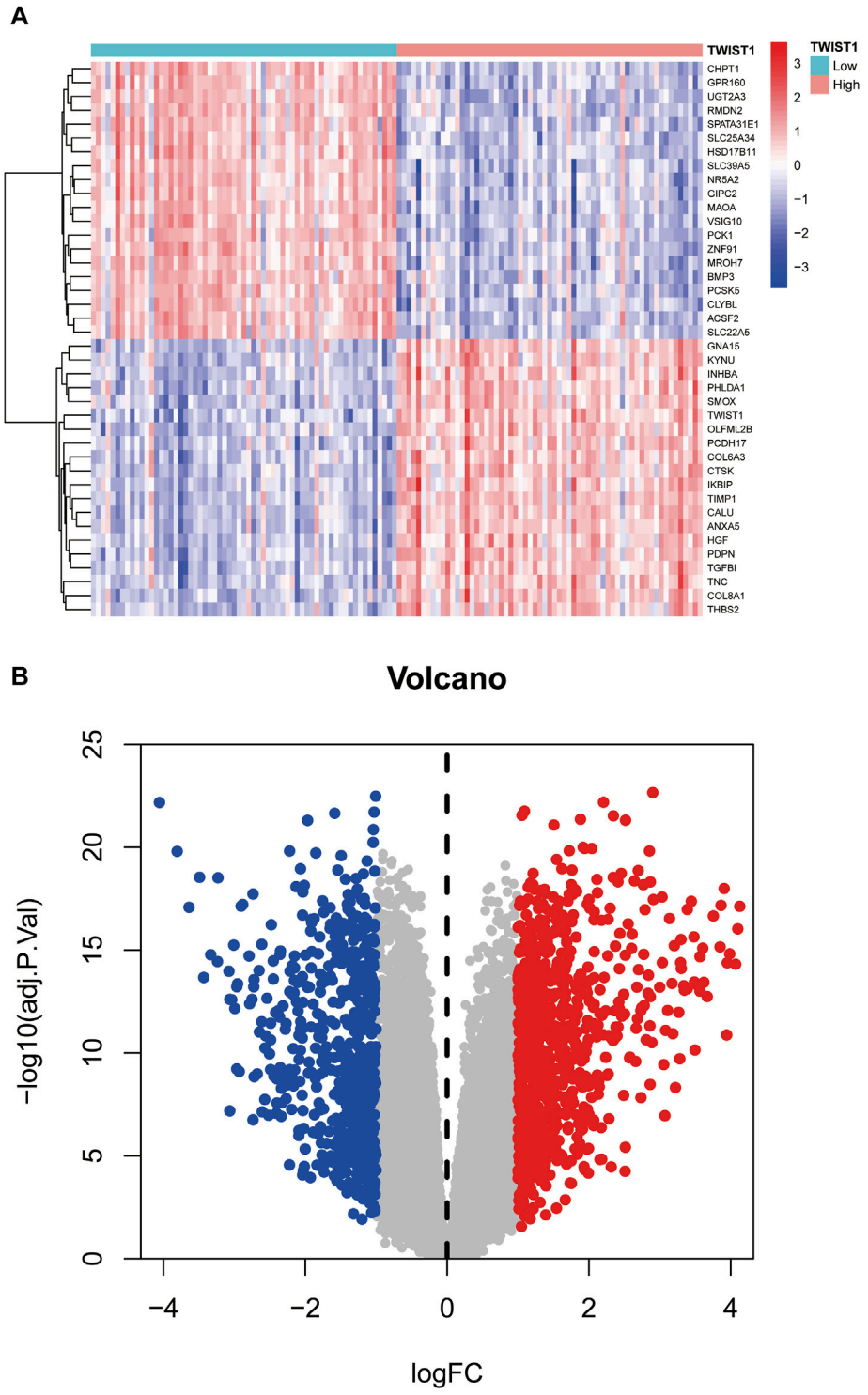


FIGURE 5
Heat map and volcano map based on the high and low expression groups of TWIST1 gene. **(A)** The heatmap of DEGs; **(B)** The volcano plots of DEGs. Red represented a high expression of DEG, while blue represented a low expression of DEG.

Employing identical methods and parameters outlined earlier, DEGs were filtered within these two groups, categorized as TWIST1 high- and low-expression groups. Analysis of DEGs between these groupings was executed via the R "Limma" package, and differential expression was visualized utilizing the R "ggplot2" package through the creation of volcano plots (p-values below 0.05 and |log2FC| exceeding 1).

2.6 Functional enrichment and gene regulatory networks analysis

Utilizing the R "clusterProfiler" package (Yu et al., 2012), Gene Ontology (GO) (Ashburner et al., 2000) analysis was conducted between elevated and diminished TWIST1 expression levels within UC samples to elucidate the implicated biological processes (BP), molecular functions (MF), and cellular components (CC). The identification of signaling pathways linked to TWIST1-associated DEGs was accomplished through Kyoto Encyclopedia of Genes and Genomes (KEGG) (Ogata et al., 1999) pathway analysis. Utilizing the "GSVA" package in R (Hänzelmann et al., 2013), the UC dataset was transformed into a gene set expression matrix. Gene Set Variation Analysis (GSVA) was then employed to meticulously examine the variations in GO and KEGG enrichment between the TWIST1 high-expression and low-expression groups. Notably, to achieve significant enrichment, the |t| value was mandated to exceed 5 for the Hallmark genome.

Adhering to the ceRNA hypothesis, predictions of TWIST1-bound miRNAs were derived using the TargetScan database (https://www.targetscan.org/vert_80/), miRDB database (http://www.mirdb.org/), and the Miranda database (https://cbio.mskcc.org/miRNA2003/miranda. html). Simultaneously, the spongeScan database (https://bioinformaticshome. com/index. html) was employed for the prediction of associated lncRNAs. The resultant networks were subsequently fine-tuned and visually represented using Cytoscape 3.92 software (Otasek et al., 2019).

2.7 Immune cell infiltrates and correlation between TWIST1

The quantification of 22 immune cell types within UC samples was accomplished through the utilization of the "CIBERSORT" software package (Newman et al., 2015). For a more in-depth analysis, exclusively data with a CIBERSORT value below $p < 0.05$ were retained. This selective process yielded a matrix detailing the fractions of immune cells present. The evaluation of immune infiltration disparity between the TWIST1 high expression and low expression groups was conducted via the Wilcoxon rank sum test. Employing the "boxplot" function within the R software

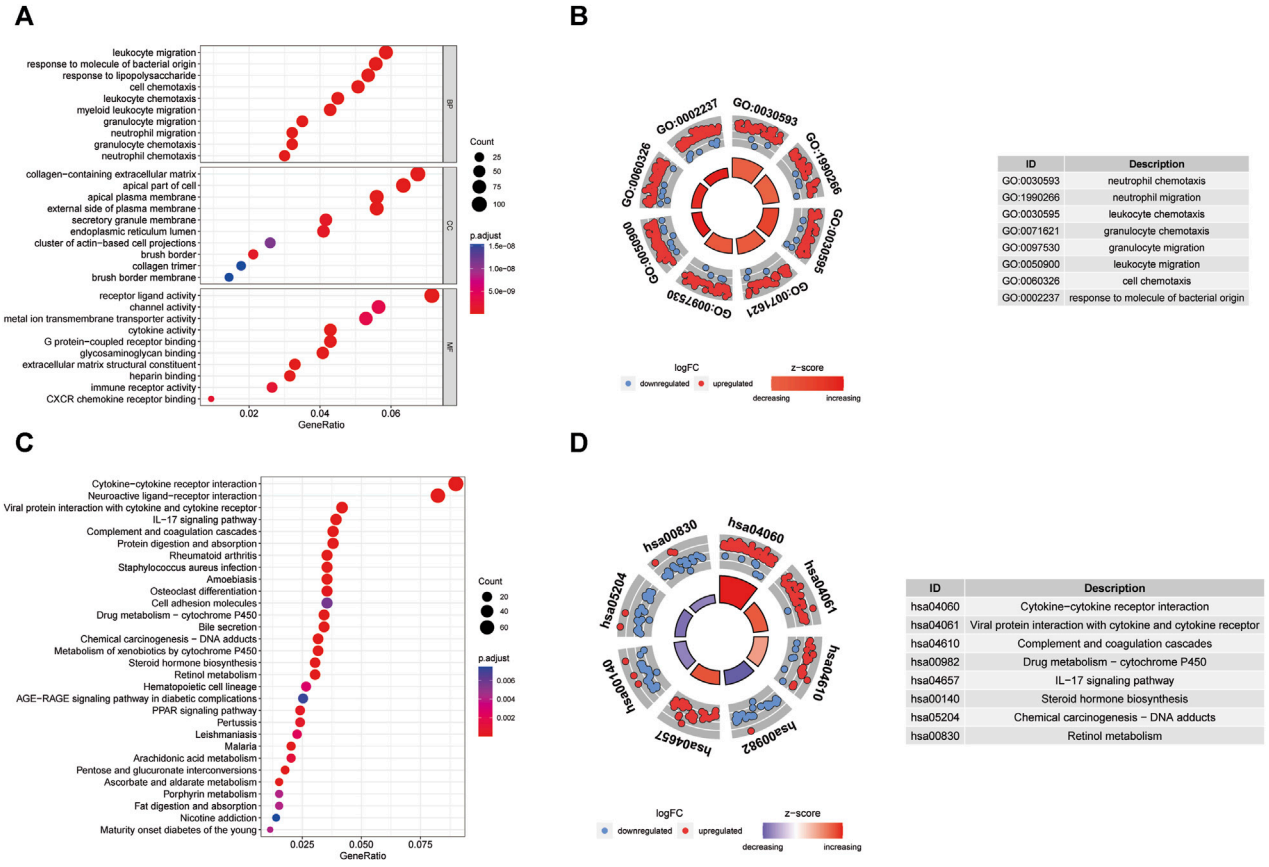


FIGURE 6
GO enrichment and KEGG analysis of DEGs. **(A)** Bubble plot of enriched GO terms. **(B)** Circos diagram of enriched GO terms. **(C)** Bubble plot of enriched KEGG terms. **(D)** Circos diagram of enriched GO terms. BP, biological process; CC, cellular component; MF, molecular function.

package, we visually depicted the contrast in immune cell infiltration levels between the two TWIST1 expression groups.

Moreover, we conducted Spearman correlation analysis to investigate the potential linkage between TWIST1 expression and immune cell infiltration. For visualization purposes, the R ggplot2 package was harnessed, allowing us to graphically represent these associations.

3 Results

3.1 Findings from genes exhibiting differential expression

The schematic portrayal of our study's methodology is depicted (Figure 1). Inclusion comprised of 461 samples obtained from healthy human intestinal mucosal biopsies (con group) and 126 samples from patients with active colonic mucosal tissue affected by UC (treat group), all sourced from the GSE193677 datasets. A comprehensive screening yielded a tally of 530 DEGs, encompassing 341 genes exhibiting upregulation and 189 genes manifesting downregulation, as highlighted (Figures 2A, B).

3.2 Machine learning to screen potential biomarkers and its diagnostic value

The LASSO logistic regression method pinpointed 87 genes as potential UC biomarkers (Figures 3A, B). Subsequently, we selected features and identified 4 optimal UC candidate genes through SVM-RFE (Figures 3C, D). The overlap between the two algorithms yielded a set of 4 genes: S100 Calcium Binding Protein P (S100P), The G protein-coupled receptor 15 (GPR15), Twist Family BHLH Transcription Factor 1 (TWIST1), and Rho Family GTPase 1 (RND1) (Figure 3E).

TWIST1 displayed significant expression differences in both GSE193677 and GSE83687 (Figures 4A, B). ROC curves were generated using data from GSE193677 and GSE83687, revealing TWIST1's AUC to be 0.717 (95% confidence interval: 0.658–0.774) and 0.897 (95% confidence interval: 0.804–0.970) in GSE193677 and GSE83687, respectively (Figures 4C, D). In the GSE193677 dataset, a

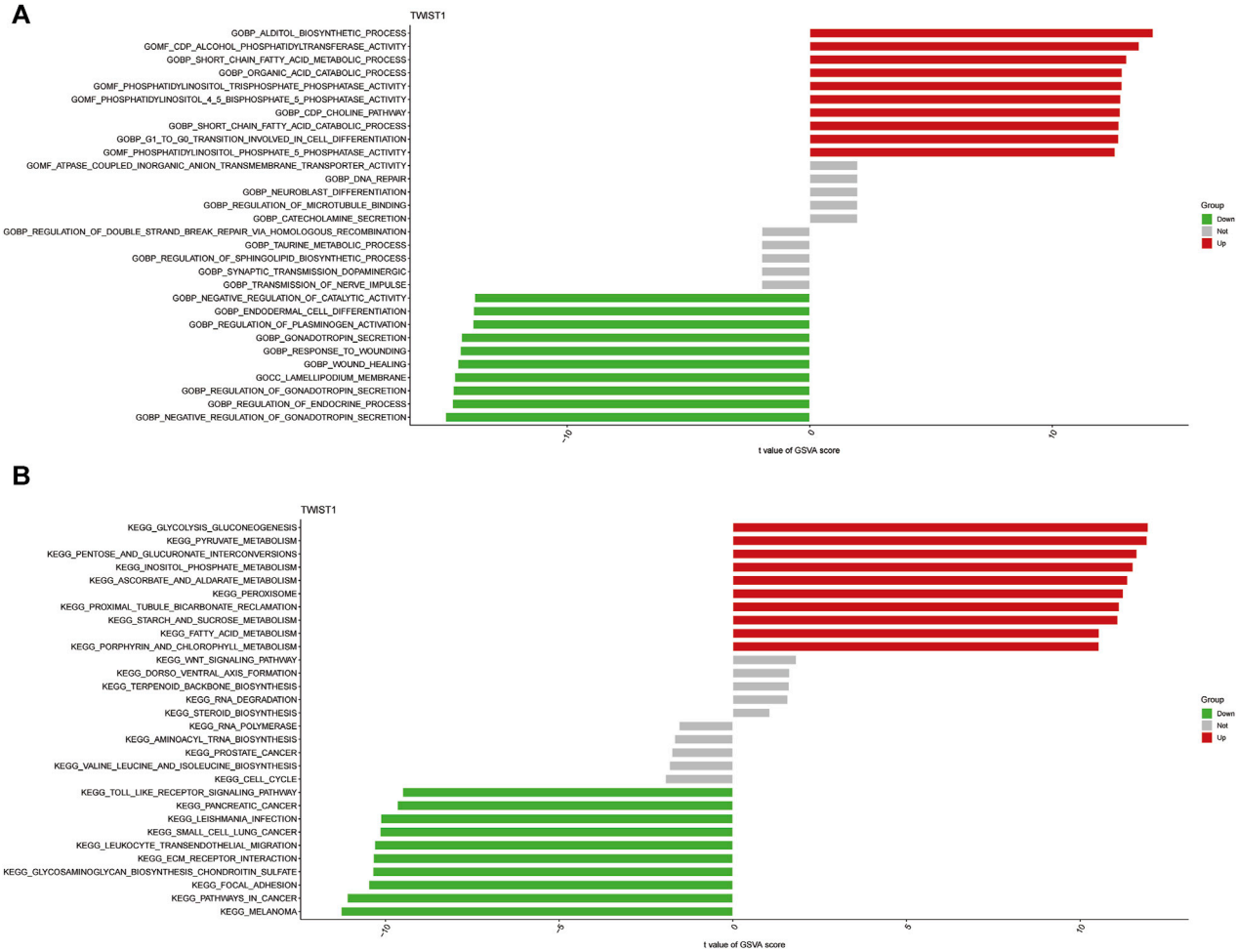


FIGURE 7
The analysis of GSVA indicates functional and pathway differences between high- and low-risk subgroups. **(A, B)** Variations in the terms of GO and KEGG pathways evaluated by GSVA between the TWIST1 high- and low-expression groups. The T values are shown using a linear model. The red column represents high enrichment in high expression groups, whereas the green column represents high enrichment in low expression groups.

significant correlation was observed between high expression levels of TWIST1 and active clinical manifestations of UC (Figure 4E).

3.3 Identification of DEGs and enrichment analysis

Within the UC sample of the GSE193677 dataset, a total of 1,518 DEGs were observed between the TWIST1 high expression and TWIST1 low expression groups, comprising 837 upregulated and 681 downregulated DEGs (Figure 5A, B).

The GO enrichment analysis demonstrated that DEGs associated with BP were predominantly linked to activities such as leukocyte migration, humoral immune response, response to molecules of bacterial origin, response to lipopolysaccharide, and cell chemotaxis. For MF, DEGs were primarily engaged in receptor-ligand activity, channel activity, metal ion transmembrane transporter activity, and cytokine activity. Concerning CC, the distribution of DEGs was prominently observed in the collagen-containing extracellular matrix, apical part of the cell, and apical plasma membrane (Figures 6A, B). The KEGG pathway enrichment analysis unveiled the enrichment of DEGs in pathways including Cytokine-cytokine receptor interaction, Viral protein interaction with cytokine and cytokine receptor, as well as Complement and IL-17 signaling (Figures 6C, D).

GSVA was conducted to further explore the terms of GO and KEGG pathways between the TWIST1 high- and low-expression groups. Top 20 upregulated terms of GO and KEGG pathways were shown (Figures 7A, B). The results of the most significant enrichment of the two groups were shown in Table 2.

3.4 CeRNA network construction of TWIST1 gene

The Supplementary Table S1 showcased the outcomes of mRNA-miRNA and lncRNA-miRNA analyses. We recognized 11 lncRNAs and 8 miRNAs, establishing their interactions through predictions and validations across databases like starBase, miRcode, Miranda, and TargetScan. The intricate interactions were graphically depicted using Cytoscape (Figure 8).

3.5 Infiltration of immune cells results

The infiltrated immune cells in different samples were analyzed using CIBERSORT and the overall relative abundances of 22 types of immune cells were shown (Figure 9A). The analysis results of infiltration degree of 22 immune cell showed that the infiltration

TABLE 2 The results of Gene Set Variation Analysis.

KEGG term	t	P-value	Sig
KEGG_MELANOMA	−11.2803	<0.001	Down
KEGG_PATHWAYS_IN_CANCER	−11.1048	<0.001	Down
KEGG_FOCAL_ADHESION	−10.4825	<0.001	Down
KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_CHONDROITIN_SULFATE	−10.3610	<0.001	Down
KEGG_ECM_RECEPTOR_INTERACTION	−10.3503	<0.001	Down
KEGG_PORPHYRIN_AND_CHLOROPHYLL_METABOLISM	10.5423	<0.001	Up
KEGG_FATTY_ACID_METABOLISM	10.5470	<0.001	Up
KEGG_STARCH_AND_SUCROSE_METABOLISM	11.0846	<0.001	Up
KEGG_PROXIMAL_TUBULE_BICARBONATE_RECLAMATION	11.1238	<0.001	Up
KEGG_PEROXISOME	11.2410	<0.001	Up
GO term	**t**	**P-value**	**Sig**
GOBP_NEGATIVE_REGULATION_OF_GONADOTROPIN_SECRETION	−15.0196	<0.001	Down
GOBP_REGULATION_OF_ENDOCRINE_PROCESS	−14.7294	<0.001	Down
GOBP_REGULATION_OF_GONADOTROPIN_SECRETION	−14.6998	<0.001	Down
GOCC_LAMELLIPODIUM_MEMBRANE	−14.6394	<0.001	Down
GOBP_WOUND_HEALING	−14.5080	<0.001	Down
GOMF_PHOSPHATIDYLINOSITOL_PHOSPHATE_5_PHOSPHATASE_ACTIVITY	12.5935	<0.001	Up
GOBP_G1_TO_G0_TRANSITION_INVOLVED_IN_CELL_DIFFERENTIATION	12.7338	<0.001	Up
GOBP_SHORT_CHAIN_FATTY_ACID_CATABOLIC_PROCESS	12.7530	<0.001	Up
GOBP_CDP_CHOLINE_PATHWAY	12.8006	<0.001	Up
GOMF_PHOSPHATIDYLINOSITOL_4_5_BISPHOSPHATE_5_PHOSPHATASE_ACTIVITY	12.8234	<0.001	Up

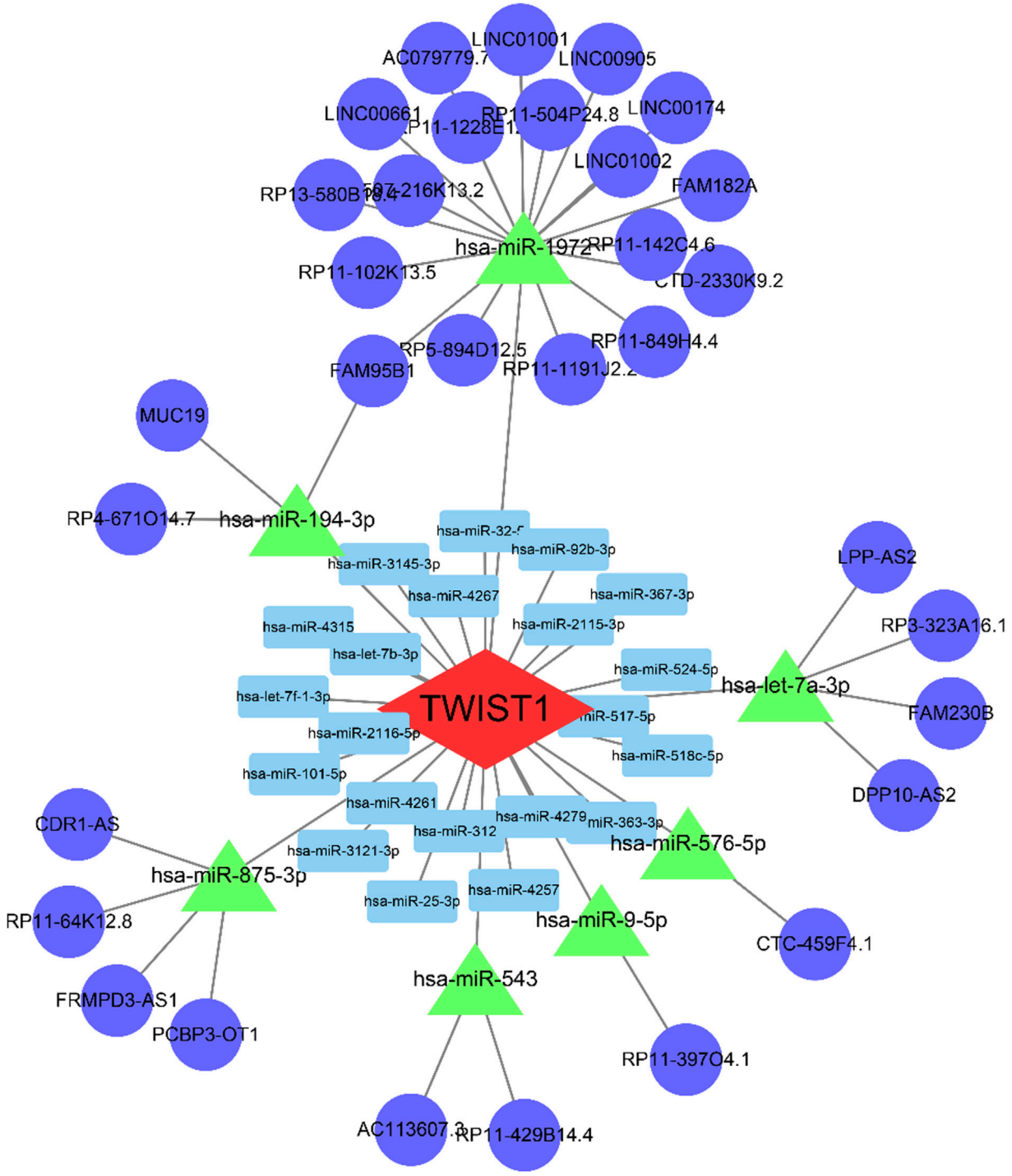


FIGURE 8
The ceRNA regulation network of 5 lncRNAs, 2 miRNAs, and 1 mRNA in patients. The blue circle indicates lncRNAs; the green rectangle indicates miRNAs; the red diamond indicates mRNA. lncRNA, long non-coding RNA; miRNA, microRNA.

of NK cells resting and Macrophages M0 was significantly different between TWIST1 high expression group and TWIST1 low expression group (Figure 9B). Further validation of the correlation study indicated that the expression level of S100A8 was correlated positively with NK cells resting and Macrophages M0 infiltration (r = 0.51, r = 0.48, all $p < 0.05$) (Figures 9C, D).

4 Discussion

UC, as one of the primary forms of IBD, is a chronic inflammatory intestinal disorder characterized by ulcers and inflammation within the intestinal tract (Kobayashi et al., 2020). The rise in Western dietary habits, improved socioeconomic status, enhanced sanitation, and advancements in vaccination have

10.3389/fgene.2024.1296570

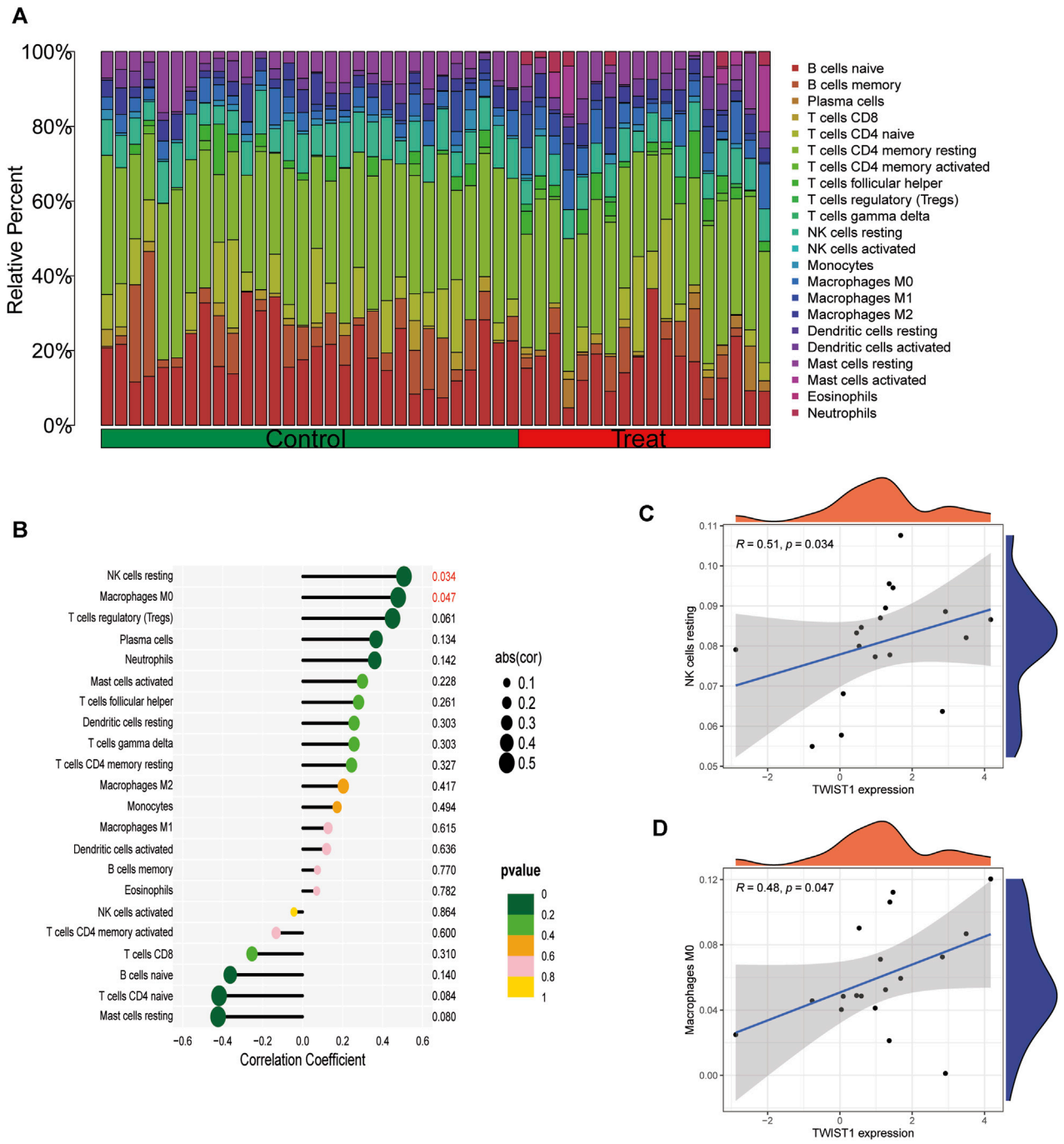


FIGURE 9
Analysis of TWIST1 and Immune Cell Infiltrates in different UC samples. **(A)** The relative percent of 22 kinds of immune cells in different lung samples. Differences in the levels of immune cells between the TWIST1 high- and low-expression groups in GSE193677. **(B)** Correlation analysis between TWIST1 and infiltrating immune cells in AS samples where red represented the positive correlation with a significant difference. **(C)** Scatter diagram indicating the correlation between TWIST1 expression and NK cells resting. **(D)** Scatter diagram indicating the correlation between TWIST1 expression and Macrophages M0 infiltration. Correlation analysis was assessed using Pearson correlation.

contributed to an increased incidence and prevalence of IBD in Asian countries (Park and Cheon, 2021). This surge is concomitant with a heightened occurrence of CAC. UC presents a significant clinical challenge, with its etiology and pathogenic mechanisms remaining largely elusive (De Souza and Fiocchi, 2016). This study aims to elucidate the potential role of the TWIST1 gene in UC through bioinformatics, machine learning, and functional enrichment analyses. Our findings offer invaluable insights into the molecular mechanisms of UC and underscore the potential of TWIST1 as a diagnostic and therapeutic target.

The Twist1 gene encodes a transcription factor encompassing a bHLH structural domain and is part of a protein family involved in organogenesis regulation (Thisse et al., 1988; Jan and January 1993; Kadesch, 1993). Recently, Twist1 has been established to play pivotal

roles not only in the development of various organs and systems but also in cancer metastasis (Yang et al., 2004; Kwok et al., 2005; Puisieux et al., 2006; Cheng et al., 2008a; Cheng et al., 2008b; Li et al., 2009; Fu et al., 2011). Studies have indicated a pronounced elevation of TWIST1 protein in tissues from UC and UC-associated colorectal cancer, with the expression intensity being greater in the latter (Anonymous, 2023). Emerging perspectives suggest that histological inflammation and its severity are among the strongest drivers of CAC risk (Shah and Itzkowitz, 2022). The bHLH transcriptional repressor - TWIST1, acting as an antagonist for NF-κB-dependent cytokine expression, partakes in the modulation of inflammation-induced immunopathology (Niesner et al., 2008; Li et al., 2009). Additionally, Twist1 may also regulate Hand proteins (Hand 1 and 2) (Firulli and Conway, 2008) and Runx2 (Rice et al., 2000; Bialek et al., 2004). These downstream targets or interacting proteins of Twist1 are known to be involved in the development of various mesenchymal derivatives and multiple physiological functions.

Existing research has demonstrated the diagnostic value of elevated TWIST1 expression in UC through immunohistochemical techniques (Anonymous, 2023). Similarly, upon acquiring high-throughput sequencing data for UC, we categorized UC samples into high and low TWIST1 expression groups. Through machine learning, lasso regression, and ROC curve analysis, we validated the diagnostic significance of elevated TWIST1 expression in UC. It is widely recognized that immune homeostasis relies on immune cells and molecules, such as innate immune cells like NK cells and macrophages M0. In the UC mucosa, metabolic abnormalities in NK cells lead to secondary infections and increased cancer risk (Zaiatz Bittencourt et al., 2021), while macrophages M0 play a role in promoting mucosal immunity and inflammatory responses in UC (Peng et al., 2023). In our study, the upregulated expression of TWIST1 in UC also increased their impact on pro-immune and pro-inflammatory cells, providing immunological support for the role of TWIST1 in the progression of UC. We also conducted GO, KEGG, and GSVA analyses on the high and low TWIST1 expression groups to explore the pathways related to TWIST1 promoting UC development.

Upon identifying TWIST1 as a biomarker, we further predicted its associated miRNA and lncRNA using databases. Notably, a study in 2022 postulated a close association between miR-9-5p and the expression of NF-κB in UC tissues (Xu et al., 2022). NF-κB plays a pivotal role in regulating immune cells and cytokines (Mantovani et al., 2004; Wang et al., 2014), and animal studies have indicated that genetic defects in the negative regulators of the canonical NF-κB pathway heighten susceptibility to colonic inflammation (Zhang et al., 2006; Vereecke et al., 2014). This regulatory axis is crucial in the onset and progression of UC. Unfortunately, other regulatory axes identified in our study have been scarcely researched in the context of UC, warranting further exploration by the scientific community.

In this study, TWIST1 is highly correlated with UC, and previous research has also indicated a strong association between TWIST1 and UC-associated colorectal cancer (Kaz et al., 2010). Given that UC serves as a precancerous lesion for UC-associated colorectal cancer (Bopanna et al., 2017), our research provides valuable clues for investigating TWIST1 as a potential risk marker in the onset, development, and transformation of UC into UC-associated colorectal cancer. This offers research directions for the future prediction and treatment of UC and UC-associated colorectal cancer. However, it is important to acknowledge certain limitations in our study. Firstly, all the data analyzed through bioinformatics methods were directly obtained from US public databases, which may not fully represent the clinical scenarios in Asian populations. Secondly, the absence of sequencing data for CAC prevented us from concurrently evaluating the diagnostic significance of TWIST1 in both UC and CAC.

Data availability statement

The data presented in this study are deposited in the GitHub repository, accessible at: https://github.com/xia-wanqiu/single_gene_analysis_for_UC.git. The sample data, GSE193677 and GSE83687, presented in this study are sourced from public databases.

Ethics statement

Ethical approval was not required for the studies involving humans because the data from the GEO database are readily accessible to the public. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

WO: Conceptualization, Methodology, Software, Validation, Visualization, Writing–original draft. ZQ: Formal Analysis, Writing–original draft. NL: Data curation, Investigation, Writing–review and editing. JZ: Data curation, Investigation, Writing–review and editing. XM: Conceptualization, Data curation, Methodology, Project administration, Writing–review and editing. YS: Investigation, Resources, Writing–review and editing. YF: Conceptualization, Data curation, Supervision, Writing–review and editing. BC: Conceptualization, Data curation, Supervision, Writing–review and editing. JH: Data curation, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Visualization, Writing–review and editing. ZY: Conceptualization, Data curation, Funding acquisition, Supervision, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The research was funded by the Jilin Provincial Department of Science and Technology under the Technical Innovation Guidance - International Scientific Collaboration, with the project number: 20230402007GH.

Acknowledgments

We extend our gratitude to the authors who shared the UC sample data in the GEO database.

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.2024.1296570/full#supplementary-material

References

Anonymous (2023). *Expression of twist1 protein in ulcerative colitis and ulcerative colitis-associated colorectal cancer and its significance-all databases[EB/OL]([no date]) [2023–09–05]*. Available at: https://www.webofscience.com/wos/alldb/full-record/CSCD:6269432.

Argmann, C., Hou, R., Ungaro, R. C., Irizar, H., Al-Taie, Z., Huang, R., et al. (2023). Biopsy and blood-based molecular biomarker of inflammation in ibd. *Gut* 72 (7), 1271–1287. doi:10.1136/gutjnl-2021-326451

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nat. Genet.* 25 (1), 25–29. doi:10.1038/75556

Bialek, P., Kern, B., Yang, X., Schrock, M., Sosic, D., Hong, N., et al. (2004). A twist code determines the onset of osteoblast differentiation. *Dev. Cell* 6 (3), 423–435. doi:10.1016/s1534-5807(04)00058-9

Bopanna, S., Ananthakrishnan, A. N., Kedia, S., Yajnik, V., and Ahuja, V. (2017). Risk of colorectal cancer in asian patients with ulcerative colitis: a systematic review and meta-analysis. *Lancet. Gastroenterology Hepatology* 2 (4), 269–276. doi:10.1016/S2468-1253(17)30004-3

Cheng, G. Z., Zhang, W., and Wang, L.-H. (2008b). Regulation of cancer cell survival, migration, and invasion by twist: akt2 comes to interplay. *Cancer Res.* 68 (4), 957–960. doi:10.1158/0008-5472.CAN-07-5067

Cheng, G. Z., Zhang, W. Z., Sun, M., Wang, Q., Coppola, D., Mansour, M., et al. (2008a). Twist is transcriptionally induced by activation of stat3 and mediates stat3 oncogenic function. *J. Biol. Chem.* 283 (21), 14665–14673. doi:10.1074/jbc.M707429200

De Souza, H. S. P., and Fiocchi, C. (2016). Immunopathogenesis of IBD: current state of the art. *Nat. Rev. Gastroenterology Hepatology* 13 (1), 13–27. doi:10.1038/nrgastro.2015.186

Du, L., and Ha, C. (2020). Epidemiology and pathogenesis of ulcerative colitis. *Gastroenterology Clin. N. Am.* 49 (4), 643–654. doi:10.1016/j.gtc.2020.07.005

Firulli, A. B., and Conway, S. J. (2008). Phosphoregulation of twist1 provides a mechanism of cell fate control. *Curr. Med. Chem.* 15 (25), 2641–2647. doi:10.2174/092986708785908987

Fu, J., Qin, L., He, T., Qin, J., Hong, J., Wong, J., et al. (2011). The twist/mi2/nurd protein complex and its essential role in cancer metastasis. *Cell Res.* 21 (2), 275–289. doi:10.1038/cr.2010.118

Ghafouri-Fard, S., Abak, A., Bahroudi, Z., Shoorei, H., Abbas Raza, S. H., and Taheri, M. (2021). The interplay between non-coding rnas and twist1 signaling contribute to human disorders. *Biomed. Pharmacother. = Biomedecine Pharmacother.* 135, 111220. doi:10.1016/j.biopha.2021.111220

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and rna-seq data. *BMC Bioinforma.* 14, 7. doi:10.1186/1471-2105-14-7

Jan, Y. N., and Jan, L. Y. (1993). HLH proteins, fly neurogenesis, and vertebrate myogenesis. *Cell* 75 (5), 827–830. doi:10.1016/0092-8674(93)90525-u

Kadesch, T. (1993). Consequences of heteromeric interactions among helix-loop-helix proteins. *Cell Growth & Differ. Mol. Biol. J. Am. Assoc. Cancer Res.* 4 (1), 49–55.

Kakiuchi, N., Yoshida, K., Uchino, M., Kihara, T., Akaki, K., Inoue, Y., et al. (2020). Frequent mutations that converge on the nfkbiz pathway in ulcerative colitis. *Nature* 577 (7789), 260–265. doi:10.1038/s41586-019-1856-1

Kaz, A., Bronner, M. P., Chen, R., Chak, A., Willis, J., and Grady, W. M. (2010). S1933 identification of novel methylated genes in barrett's esophagus with high-grade dysplasia and esophageal adenocarcinoma using methylation arrays. *GASTROENTEROLOGY* 138 (5), S283–S284. doi:10.1016/s0016-5085(10)61304-2

Kobayashi, T., Siegmund, B., Le Berre, C., Wei, S. C., Ferrante, M., Shen, B., et al. (2020). Ulcerative colitis. *Nat. Rev. Dis. Prim.* 6 (1), 74. [2023–08–19]. doi:10.1038/s41572-020-0205-x

Kumar, R., and Indrayan, A. (2011). Receiver operating characteristic (roc) curve for medical researchers. *Indian Pediatr.* 48 (4), 277–287. doi:10.1007/s13312-011-0055-4

Kwok, W. K., Ling, M.-T., Lee, T.-W., Lau, T. C. M., Zhou, C., Zhang, X., et al. (2005). Up-regulation of twist in prostate cancer and its implication as a therapeutic target. *Cancer Res.* 65 (12), 5153–5162. doi:10.1158/0008-5472.CAN-04-3785

Li, Q.-Q., Xu, J.-D., Wang, W.-J., Cao, X. X., Chen, Q., Tang, F., et al. (2009). Twist1-mediated adriamycin-induced epithelial-mesenchymal transition relates to multidrug resistance and invasive potential in breast cancer cells. *Clin. Cancer Res. Official J. Am. Assoc. Cancer Res.* 15 (8), 2657–2665. doi:10.1158/1078-0432.CCR-08-2372

Mantovani, A., Sica, A., Sozzani, S., Allavena, P., Vecchi, A., and Locati, M. (2004). The chemokine system in diverse forms of macrophage activation and polarization. *Trends Immunol.* 25 (12), 677–686. doi:10.1016/j.it.2004.09.015

Murre, C., McCaw, P. S., Vaessin, H., Caudy, M., Jan, L. Y., Jan, Y. N., et al. (1989). Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common dna sequence. *Cell* 58 (3), 537–544. doi:10.1016/0092-8674(89)90434-0

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12 (5), 453–457. doi:10.1038/nmeth.3337

Niesner, U., Albrecht, I., Janke, M., Doebis, C., Loddenkemper, C., Lexberg, M. H., et al. (2008). Autoregulation of th1-mediated inflammation by *twist1*. *J. Exp. Med.* 205 (8), 1889–1901. doi:10.1084/jem.20072468

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 27 (1), 29–34. doi:10.1093/nar/27.1.29

Otasek, D., Morris, J. H., Bouças, J., Pico, A. R., and Demchak, B. (2019). Cytoscape automation: empowering workflow-based network analysis. *Genome Biol.* 20 (1), 185. doi:10.1186/s13059-019-1758-4

Park, J., and Cheon, J. H. (2021). Incidence and prevalence of inflammatory bowel disease across asia. *Yonsei Med. J.* 62 (2), 99–108. doi:10.3349/ymj.2021.62.2.99

Peng, Y., Cui, Z., Li, Y., Yang, Y., Liu, R., Luan, Y., et al. (2023). Characteristics of immune cell infiltration in inflamed mucosa of ulcerative colitis patients, hub gene candidates and key pathways in intestinal macrophage expression. *ALL LIFE* 16 (1), 2229042. doi:10.1080/26895293.2023.2229042

Peters, L. A., Perrigoue, J., Mortha, A., Iuga, A., Song, W. M., Neiman, E. M., et al. (2017). A functional genomics predictive network model identifies regulators of inflammatory bowel disease. *Nat. Genet.* 49 (10), 1437–1449. doi:10.1038/ng.3947

Puisieux, A., Valsesia-Wittmann, S., and Ansieau, S. (2006). A twist for survival and cancer progression. *Br. J. Cancer* 94 (1), 13–17. doi:10.1038/sj.bjc.6602876

Ren, H., Du, P., Ge, Z., Jin, Y., Ding, D., Liu, X., et al. (2016). TWIST1 and bmi1 in cancer metastasis and chemoresistance. *J. Cancer* 7 (9), 1074–1080. doi:10.7150/jca.14031

Rice, D. P., Aberg, T., Chan, Y., Tang, Z., Kettunen, P. J., Pakarinen, L., et al. (2000). Integration of fgf and twist in calvarial bone and suture development. *Dev. Camb. Engl.* 127 (9), 1845–1855. doi:10.1242/dev.127.9.1845

Shah, S. C., and Itzkowitz, S. H. (2022). Colorectal cancer in inflammatory bowel disease: mechanisms and management. *Gastroenterology* 162 (3), 715–730.e3. doi:10.1053/j.gastro.2021.10.035

Suykens, J. A. K., and Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Process. Lett.* 9 (3), 293–300. doi:10.1023/A:1018628609742

Thisse, B., Stoetzel, C., Gorostiza-Thisse, C., and Perrin-Schmitt, F. (1988). Sequence of the twist gene and nuclear localization of its protein in endomesodermal cells of early drosophila embryos. *EMBO J.* 7 (7), 2175–2183. doi:10.1002/j.1460-2075.1988.tb03056.x

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Methodol.* 58 (1), 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x

Tong, L., Hao, H., Zhang, Z., Lv, Y., Liang, X., Liu, Q., et al. (2021). Milk-derived extracellular vesicles alleviate ulcerative colitis by regulating the gut immunity and reshaping the gut microbiota. *Theranostics* 11 (17), 8570–8586. doi:10.7150/thno.62046

Vereecke, L., Vieira-Silva, S., Billiet, T., van Es, J. H., Mc Guire, C., Slowicka, K., et al. (2014). A20 controls intestinal homeostasis through cell-specific activities. *Nat. Commun.* 5, 5103. doi:10.1038/ncomms6103

Wang, N., Liang, H., and Zen, K. (2014). Molecular mechanisms that influence the macrophage m1–m2 polarization balance. *Front. Immunol.* 5, 614. doi:10.3389/fimmu.2014.00614

Xu, M., Kong, Y., Chen, N., Peng, W., Zi, R., Jiang, M., et al. (2022). Identification of immune-related gene signature and prediction of cerna network in active ulcerative colitis. *Front. Immunol.* 13, 855645. doi:10.3389/fimmu.2022.855645

Yang, J., Mani, S. A., Donaher, J. L., Ramaswamy, S., Itzykson, R. A., Come, C., et al. (2004). Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* 117 (7), 927–939. doi:10.1016/j.cell.2004.06.006

Yashiro, M. (2014). Ulcerative colitis-associated colorectal cancer. *World J. Gastroenterology* 20 (44), 16389–16397. doi:10.3748/wjg.v20.i44.16389

Yu, G., Wang, L.-G., Han, Y., and He, Q. Y. (2012). ClusterProfiler: an r package for comparing biological themes among gene clusters. *Omics A J. Integr. Biol.* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Zaiatz Bittencourt, V., Jones, F., Tosetto, M., Doherty, G. A., and Ryan, E. J. (2021). Dysregulation of metabolic pathways in circulating natural killer cells isolated from inflammatory bowel disease patients. *J. Crohn's Colitis* 15 (8), 1316–1325. doi:10.1093/ecco-jcc/jjab014

Zhang, J., Stirling, B., Temmerman, S. T., Ma, C. A., Fuss, I. J., Derry, J. M. J., et al. (2006). Impaired regulation of NF-kappaB and increased susceptibility to colitis-associated tumorigenesis in CYLD-deficient mice. *J. Clin. Investigation* 116 (11), 3042–3049. doi:10.1172/JCI28746

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 18 April 2024
DOI 10.3389/fgene.2024.1242974

OPEN ACCESS

EDITED BY
Che Kang Lim,
Singapore General Hospital, Singapore

REVIEWED BY
Apt Lalu Muhammad Irham,
Ahmad Dahlan University, Indonesia
Qiaoyang Sun,
National Neuroscience Institute (NNI), Singapore

*CORRESPONDENCE
Qi Wang,
wangqi710@126.com
Ji Wang,
doctorwang2009@126.com

†These authors have contributed equally to this work

RECEIVED 20 June 2023
ACCEPTED 25 March 2024
PUBLISHED 18 April 2024

CITATION
Sun P, Wang Y, Liu X, Li Z, Cui D, Li Q, Wang Q and Wang J (2024), Methylation-driven mechanisms of allergic rhinitis during pollen and non-pollen seasons using integrated bioinformatics analysis.
Front. Genet. 15:1242974.
doi: 10.3389/fgene.2024.1242974

Methylation-driven mechanisms of allergic rhinitis during pollen and non-pollen seasons using integrated bioinformatics analysis

Pengcheng Sun[1,2†], Yi Wang[1,2†], Xing Liu[1,2], Zhuqing Li[1,2], Diankun Cui[1,2], Qianru Li[3], Qi Wang[2*†] and Ji Wang[2*†]

[1]College of Traditional Chinese Medicine, Beijing University of Chinese Medicine, Beijing, China, [2]National Institute of Traditional Chinese Medicine Constitution and Preventive Medicine, Beijing University of Chinese Medicine, Beijing, China, [3]Qinghai Golmud Jianqiao Hospital, Golmud, Qinghai, China

Background: Allergic rhinitis (AR) is a widespread allergic airway disease that results from a complex interplay between genetic and environmental factors and affects approximately 10%–40% of the global population. Pollen is a common allergen, and exposure to pollen can cause epigenetic changes. However, the mechanism underlying pollen-induced DNA methylation changes and their potential effects on the allergic march are still unclear. The purpose of this study was to explore the methylation-driven mechanisms of AR during the pollen and non-pollen seasons using bioinformatics analysis and to investigate their relationship with asthma.

Methods: We downloaded DNA methylation and gene expression data from the GEO database (GSE50387: GSE50222, GSE50101) and identified differentially methylated positions (DMPs) and differentially expressed genes (DEGs) during the pollen and non-pollen seasons using the CHAMP and limma packages. Through correlation analysis, we identified methylation-driven genes and performed pathway enrichment analysis to annotate their functions. We incorporated external data on AR combined with asthma (GSE101720) for analysis to identify key CpGs that promote the transformation of AR to asthma. We also utilized external data on olive pollen allergy (GSE54522) for analysis to validate the methylation-driven genes. Weighted correlation network analysis (WGCNA) was employed to identify gene modules significantly correlated with pollen allergy. We extracted genes related to the key methylation-driven gene *ZNF667-AS1* from the significant module and performed pathway intelligent clustering using KOBAS-i. We also utilized gene set enrichment analysis to explore the potential function of *ZNF667-AS1*.

Results: We identified 20 and 24 CpG-Gene pairings during the pollen and non-pollen seasons. After incorporating external data from GSE101720, we found that *ZNF667-AS1* is a key gene that may facilitate the transformation of AR into asthma during the pollen season. This finding was further validated in another external dataset, GSE54522, which is associated with pollen allergy. WGCNA identified 17 modules, among which the blue module showed significant correlation with allergies. *ZNF667-AS1* was located in the blue module. We performed pathway analysis on the genes correlated with *ZNF667-AS1* extracted from the blue module and identified a prominent cluster of pathways in the KOBAS-i results, including Toll-like receptor (*TLR*) family, *MyD88*, *MAPK*, and oxidative stress.

Gene set enrichment analysis around *cg05508084* (paired with *ZNF667-AS1*) also indicated its potential involvement in initiating and modulating allergic inflammation from the perspective of *TLR* and *MAPK* signaling.

Conclusion: We identified methylation-driven genes and their related pathways during the pollen and non-pollen seasons in patients with AR and identified key CpGs that promote the transformation of AR into asthma due to pollen exposure. This study provides new insights into the underlying molecular mechanisms of the transformation of AR to asthma.

KEYWORDS

allergic rhinitis, asthma, pollen season, gene expression profile, DNA methylation profile, bioinformatics

Introduction

Allergic diseases, including allergic rhinitis (AR), asthma, and atopic dermatitis, represent significant health concerns in global populations (Paller et al., 2019; Miller et al., 2021; Ständer, 2021; Zhang et al., 2021). There is a close association between allergic diseases, and the development of AR may be related to the manifestation of systemic allergy, including asthma (Bousquet et al., 2020b; Dharmage et al., 2022). Thirty percent of individuals with rhinitis also have asthma, and over eighty percent of individuals with asthma have symptoms of rhinitis (Nappi et al., 2022).

Pollen is a widespread aeroallergen that can induce allergic disease (Taylor et al., 2007; Schutzmeier et al., 2022). For now, it is recognized that more than 150 pollen allergens originating from grasses, trees, and weeds contribute significantly to an allergic response (Xie et al., 2019). Worldwide, the sensitization rate to pollen allergens is approximately 40% (Pointner et al., 2020). According to the International Study of Asthma and Allergies in Childhood (ISAAC), the prevalence of pollen sensitization in children increases by approximately 0.3% per year (Suanno et al., 2021). Approximately 20% of the US population suffers from pollen allergies, while about 20% of the population in Europe is affected by grass pollen allergies (García-Mozo, 2017). About 18.5% of the population in the northern grasslands of China is affected by pollen-induced AR (Wang et al., 2018). The incidence of pollen allergies exhibits geographic variation, influenced by bioclimatic conditions and the distribution of allergenic plants. Several studies have demonstrated that the exposure to pollen can induce DNA methylation (one of the epigenetic control of gene expression) for AR patients (North et al., 2018; Watanabe et al., 2021; Yang et al., 2022). The underlying reason is that environmental exposures can influence DNA methylation, which mediates the interaction between the environment and genotype to impact clinical phenotype (Law and Holland, 2019). Concerning the relationship between methylation and AR, epigenome-wide association studies (EWAS) have found specific CpG sites in AR patients (Qi et al., 2020). DNA methylation may help distinguish allergic patients from healthy individuals (Choi et al., 2021). According to the sensitization on cyclic pollens or year-round allergens, AR can be classified as seasonal or perennial (Greiner et al., 2011). For seasonal allergic rhinitis (SAR) patients, a previous study demonstrated that DNA methylation profiles instead of gene expression profiles can clearly and robustly distinguish them from healthy controls, during the pollen and non-pollen seasons (Nestor et al., 2014). However, how pollen-induced DNA methylation affects SAR patients by changing gene expression has not been fully elucidated.

To investigate the role of DNA methylation in the regulation of gene expression in SAR, we downloaded publicly available data and performed a correlation analysis of significant differentially methylated positions (DMPs) with critical differentially expressed genes (DEGs). Our bioinformatics analysis aimed to identify potential epigenetic mechanisms that contribute to the development and progression of SAR. To support our findings and investigate the role pollen plays in the evolution of SAR to asthma, we incorporated external data on the co-morbidity of AR and asthma and compared it with the analysis results of SAR during the pollen and non-pollen seasons. The flowchart of the analysis is shown in Figure 1.

Materials and methods

DNA methylation and gene expression data resources

The microarray datasets GSE50222 (Last update date: 22 Mar 2019) and GSE50101 (Last update date: 13 Aug 2018) were downloaded from the GEO database (http://www.ncbi.nlm.nih.gov/geo/) (Edgar et al., 2002; Barrett et al., 2013). They were derived from a super series GSE50387 (Nestor et al., 2014) that contains methylation chip and gene chip data of $CD4^+$ T cells from the same group of AR patients who were recruited from Sweden people in Europe. Positive skin prick test results and/or positive ImmunoCAP Rapid results to birch and/or grass pollen were the two main indicators of SAR, which were both negative in health controls. Patients with perennial symptoms or asthma were not included. GSE50222 is a methylation dataset that contains 32 samples, which are divided into four groups: AR of the pollen season (n = 8), AR of the non-pollen season (n = 8), healthy controls (HCs) of the pollen season (n = 8), and HCs of the non-pollen season (n = 8). The chip platform used was GPL13534 (Illumina HumanMethylation450 BeadChip). GSE50101 is a gene expression dataset containing 38 samples, which are divided into four groups: AR of the pollen season (n = 9), AR of the non-pollen season (n = 9), HCs of the pollen season (n = 10), and HCs of the non-pollen season (n = 10). The chip platform used was GPL10558 (Illumina HumanHT-12 V4.0 Expression Beadchip).

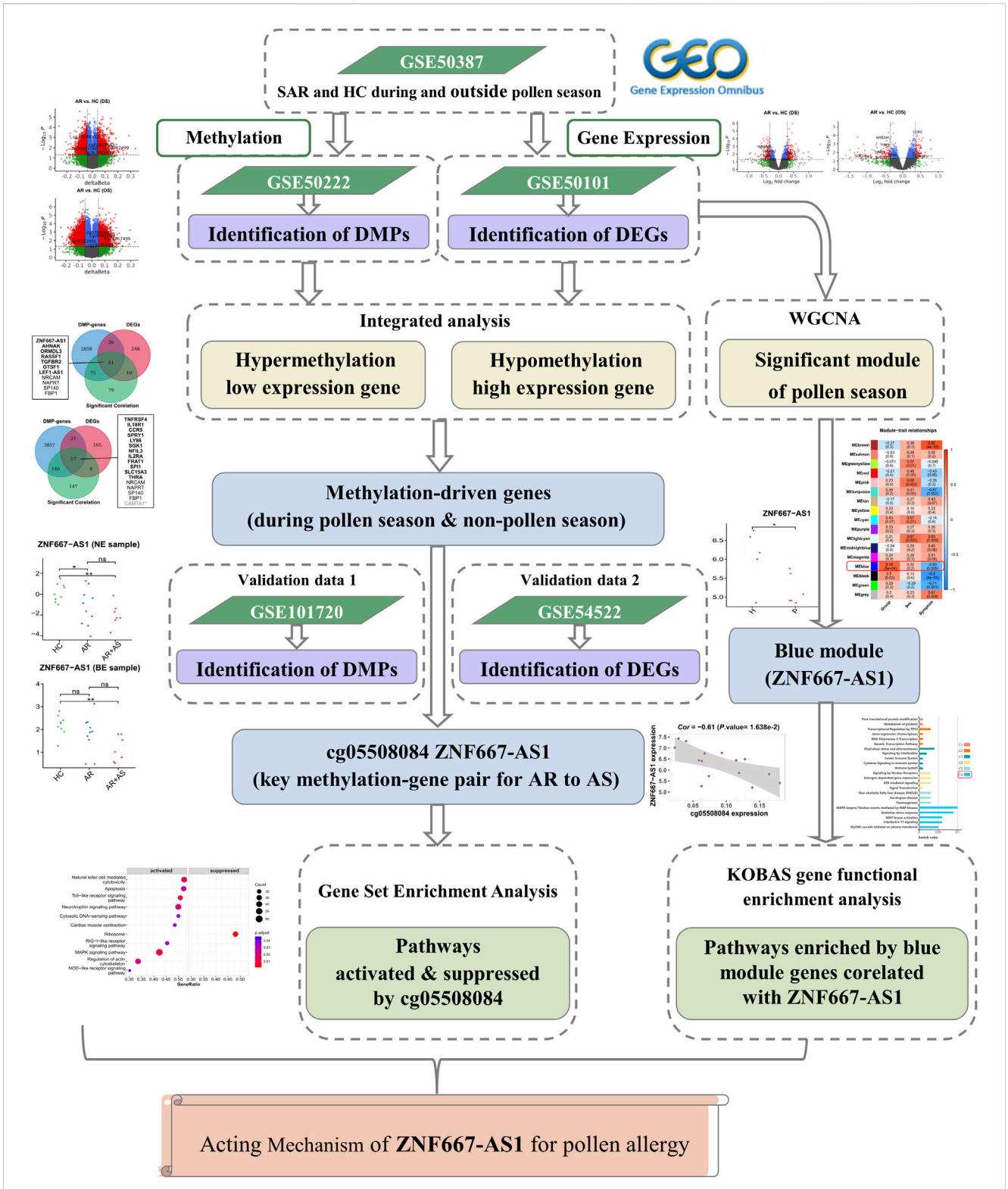


FIGURE 1
The flowchart of the analysis. DMP: differentially methylated position; DEG: differentially expressed gene; WGCNA: weighted correlation network analysis; AR: allergic rhinitis; SAR: seasonal allergic rhinitis; AS: asthma; HC: healthy control.

Data preprocessing and DMP screening

R (version 4.2.3) (R Core Team, 2023) was used for data preprocessing. First, we determined the missing rates for both the probes and the samples, and we kept only the probes and samples with missing rates under 10%. We eliminated 'bad' probes like cross-reactive/non-specific probes and BOWTIE2 multi-mapped probes (Chen et al., 2013; Pidsley et al., 2016). Afterward, the beta matrix was calculated

using the following equation: β = M/(M + U+100). For our samples, we extracted the grouping information using the GEOquery package (Davis and Meltzer, 2007). To further filter out the probes, we used the ChAMP package (version 2.24.0) (Morris et al., 2014; Tian et al., 2017) and its champ.filter function (Wang et al., 2023).

We conducted a differentially methylated positions (DMP) analysis between the AR group and the HC group, both during the pollen and non-pollen seasons. Taking into account the effect of Type I and Type II probes, we first conducted quality control using champ.QC function and then applied the champ.norm function for standardization. Additionally, we utilized champ.SVD function to look at the principal components and champ.DMP to find DMPs.

Firstly, we conducted quality control using champ.QC function and then applied champ.norm function for standardization to correct for the effects of Type-I and Type-II probes. We also used champ.SVD function to examine principal components and identified DMPs using champ.DMP function. The following criteria were set to further screen for methylation-driven sites (Guo et al., 2020):

(1) $\mathrm{abs}(\Delta\beta) > \mathrm{mean}(\mathrm{abs}(\Delta\beta)) + 2*\mathrm{sd}(\mathrm{abs}(\Delta\beta))$
(2) p-value < 0.05
(3) The location belongs to the promoter regions (1stExon, 5′UTR, TSS1500, TSS200).

Data preprocessing and DEG screening

We conducted a DEG analysis between the AR group and the HC group, both during the pollen and non-pollen seasons.

First, we determined the missing rate for both the samples and the probes, keeping only the probes with a missing rate of less than 10%. According to detection p-values, we performed background correction on the samples using the nec function in the limma package (version 3.50.3) (Ritchie et al., 2015). The second step involved probe filtration. The raw gene expression matrix was split by group. Significant probes were defined as those with p-value < 0.05 in at least 50% of the samples. For non-significant probes, their expression values were replaced with the average expression value of these probes across all samples within each group. Third, we used limma package to do DEG analysis with filtering criteria set to $\mathrm{abs}(\log_2\mathrm{FC}) > \mathrm{mean}(\mathrm{abs}(\log_2\mathrm{FC})) + 2*\mathrm{sd}(\mathrm{abs}(\log_2\mathrm{FC}))$ and p-value < 0.05. Finally, the probes were annotated and the duplicate genes were removed.

We create a Venn diagram with the R package VennDiagram (v1.73) (Chen, 2018) to depict the link between the upregulated and downregulated DEGs discovered during the pollen and non-pollen seasons. Then, we selected the unique parts of these four groups of genes and conducted gene function enrichment analysis for each group with Kyoto Encyclopedia of Genes and Genomes database (KEGG) (Kanehisa and Goto, 2000; Kanehisa et al., 2021) using clusterProfiler package (version 4.7.1.003) (Yu et al., 2012; Wu et al., 2021) and org.Hs.eg.db package (version 3.16.0) (Carlson, 2021).

Integrated analysis of DNA methylation and gene expression data

We matched the samples between methylation and gene chips according to the patient ID, and obtained the correspondence between CpG sites and genes through annotation information. Then we calculated the Spearman correlation coefficient between the methylation β values at CpG sites and the gene expression values. CpG-Gene pairs were filtered with p-value < 0.05 and $Cor < -0.4$. Only the DMP-DEG pairs were retained. Finally, we compared the results of DMP related genes, DEGs, and the genes with significant correlation coefficients using a Venn diagram both in the pollen season and the non-pollen season.

We conducted gene function enrichment analysis on the methylation-driven genes obtained during the pollen and non-pollen seasons based on the hypergeometric distribution. The pathway databases used included KEGG database (Kanehisa and Goto, 2000; Kanehisa et al., 2021), Reactome database (Gillespie et al., 2022), and Gene Ontology (GO) database (Ashburner et al., 2000; Gene Ontology Consortium, 2021) limited to biological processes (BP). Enrichment analysis was performed using the enrichKEGG function (Yu et al., 2012; Wu et al., 2021), enrichPathway function (Yu and He, 2016), and enrichGO function (Yu et al., 2012; Wu et al., 2021) for KEGG, Reactome, and GO, respectively. Finally, the results were compared and visualized.

Comparison of methylation-driven genes with data on allergic comorbidity

The allergic comorbidity sequencing dataset GSE101720 (Last update date: 17 Dec 2020) was also downloaded from the GEO database. It includes samples from people who have AR and asthma, people who merely have AR and healthy people (Giovannini-Chami et al., 2018). The dataset contains nasal epithelial samples (n = 26, AR with asthma = 7, AR = 10, HC = 9) and bronchial epithelial samples (n = 26, AR with asthma = 7, AR = 10, HC = 9) that were sequenced on the Illumina NextSeq 500 platform (GPL18573).

With samples from the nasal epithelium and bronchial epithelium, we used the limma package to identify DEGs between the AR with asthma group and the HC group, as well as between the AR group and the HC group. The filtering criterion (Luo et al., 2019) for DEGs was set as p-value < 0.05 and $\mathrm{abs}(\log_2\mathrm{FC}) > (\mathrm{mean}(\mathrm{abs}(\log_2\mathrm{FC})) + 2*\mathrm{sd}(\mathrm{abs}(\log_2\mathrm{FC})))$.

We retrieved the methylation-driven genes from the previous section and compared them to the DEGs from this part.

Comparison of methylation-driven genes with data on olive pollen allergy

The data on olive pollen allergy, GSE54522 (Last update date: 25 Mar 2019), was downloaded from the GEO database. This dataset includes peripheral blood mononuclear cells (PBMCs) from 6 olive pollen-allergic patients and 6 HC subjects, which were stimulated with olive pollen for 24 h (Calzada et al., 2015). The olive pollen-allergic patients met the following criteria: seasonal rhinitis and/or asthma from April to June, a positive skin prick test result for *O. europaea* pollen extract, and no previous immunotherapy. From this dataset, we selected samples stimulated with pollen extract during the pollen season (n = 6) and compared them to the stimulated samples from the HC group (n = 5). The filtering criterion

(Luo et al., 2019) for identifying DEGs was set as a p-value < 0.05 and abs($\log_2$FC) > mean(abs($\log_2$FC)) + 2*sd(abs($\log_2$FC)). We then compared the methylation-driven genes with the DEGs identified in this analysis.

Weighted correlation network analysis for GSE50101

We utilized Weighted Correlation Network Analysis (WGCNA) (Langfelder and Horvath, 2008) to analyze the gene expression data from GSE50101, aiming to explore the gene expression patterns in samples during the pollen season. To ensure data heterogeneity and analysis accuracy, the following gene filtering steps were applied: a) Genes were tested using the goodGenes function; b) Genes with an upper 25% median absolute deviation were selected (Wei et al., 2020). For sample selection in WGCNA, the following steps were performed: a) Samples were tested using the goodSamples function; b) A sample clustering plot was generated using the hclust function, and outlier samples were removed. The prepared gene expression matrix was used to calculate the Soft Threshold (with R^2 cutoff set to the default value of 0.85). Considering the involvement of numerous genes in allergies, the minModuleSize was set to 100 to construct a weighted gene co-expression network. For the obtained gene modules, module eigengenes were calculated and used to assess the correlation with traits (grouping, gender, symptom scores). Key modules were selected based on a correlation threshold of *Cor* > 0.4 and a p-value < 0.05. Within the key module, the correlation between all genes and *ZNF667-AS1* expression was calculated, and genes with a correlation of *Cor* > 0.4 and a p-value < 0.05 were selected. Pathway analysis was performed using KOBAS-i (Bu et al., 2021), with pathway databases including KEGG, Reactome, BioCyc, and PANTHER. Pathways with a Corrected p-value < 0.05 were further subjected to intelligent clustering and visualization.

Gene set enrichment analysis for *cg05508084*

ZNF667-AS1 was identified as a key methylation-driven gene during the pollen season, and it is regulated by *cg05508084*. To explore the potential function of *ZNF667-AS1*, we calculate the association between *cg05508084* and the levels of gene expression for the pollen season. Following the results of the correlation analysis, single-gene Gene Set Enrichment Analysis (GSEA) was conducted with the databases KEGG and Reactome from MSigDB (Subramanian et al., 2005) using the msigdbr package (version 7.5.1) (Dolgalev, 2022). Visualization was done of the pathways that *cg05508084* either activated or inhibited.

Results

Identification of DMPs during the pollen and non-pollen seasons

The general workflow for analyzing Illumina HumanMethylation450 BeadChip data includes quality control, normalization, differential expression analysis, and annotation. Our quality control consists of three steps: a) filtering based on probe and sample missing rate, removing probes and samples with missing rates exceeding 10%; b) filtering based on unique probe annotation, removing 'bad' probes with non-unique annotations; c) filtering based on Methylated matrix, UnMethylated matrix, and Detected p-value. The study began by filtering out methylation probes that had a missing rate of over 10%. This resulted in 4,49,506 probes remaining out of the initial 4,85,577. Next, non-specific probes were filtered out, leaving 4,13,172 methylation probes. The probes were further filtered using champ.filter, with 3,89,332 remaining in the end. The beta matrix was standardized between type I and type II probes using champ.norm. DMP analysis was performed using champ.DMP, resulting in the identification of 11,275 DMPs during pollen season and 16,975 DMPs during the non-pollen season. In champ.DMP results, DMPs are annotated with genomic locations and corresponding genes. Of these, 2,987 DMPs of the pollen season and 4,458 DMPs of the non-pollen season were located in the promoter region. Volcano plots for DMPs are displayed in Figures 2A, B, while Figures 2C, D compare the distribution of DMPs of the pollen and the non-pollen seasons. The principal component analysis (PCA) plot and heatmap of the methylation data (GSE50222) are shown in Supplementary Figure S1.

Identification of DEGs during the pollen and non-pollen seasons

The analysis began by filtering out gene probes with a missing rate of over 10%, which resulted in 47,314 probes remaining out of the initial 47,323. In the AR group, 18,327 and 17,848 significant probes were retained during the pollen and non-pollen seasons. In the HC group, 18,591 and 18,371 significant probes were retained during the pollen and non-pollen seasons. Using $\log_2$FC and p-value criteria, a total of 293 DEGs were selected for the pollen season, out of which 160 were upregulated, and 133 were downregulated in AR. Similarly, 212 DEGs were selected for the non-pollen season, among which 95 were upregulated, and 117 were downregulated in AR. The volcano plots are shown in Figure 3A, B. The comparison of upregulated and downregulated DEGs during the pollen and non-pollen seasons can be seen in Figure 3C, D, the KEGG pathways enriched by the distinct genes in each of the four gene sets are contrasted. The PCA plot and heatmap of the gene expression data (GSE50101) are shown in Supplementary Figure S2.

Results of integrated analysis of DNA methylation and gene expression data

We identified 15 relevant pairs of samples by matching the sample IDs from the methylation (GSE50222) and gene expression data (GSE50101) during the pollen and non-pollen seasons. Subsequently, we determined the correlation coefficient between CpG sites and genes using consistent samples. A total of 262 and 470 CpG-Gene pairings were found during the pollen and non-pollen

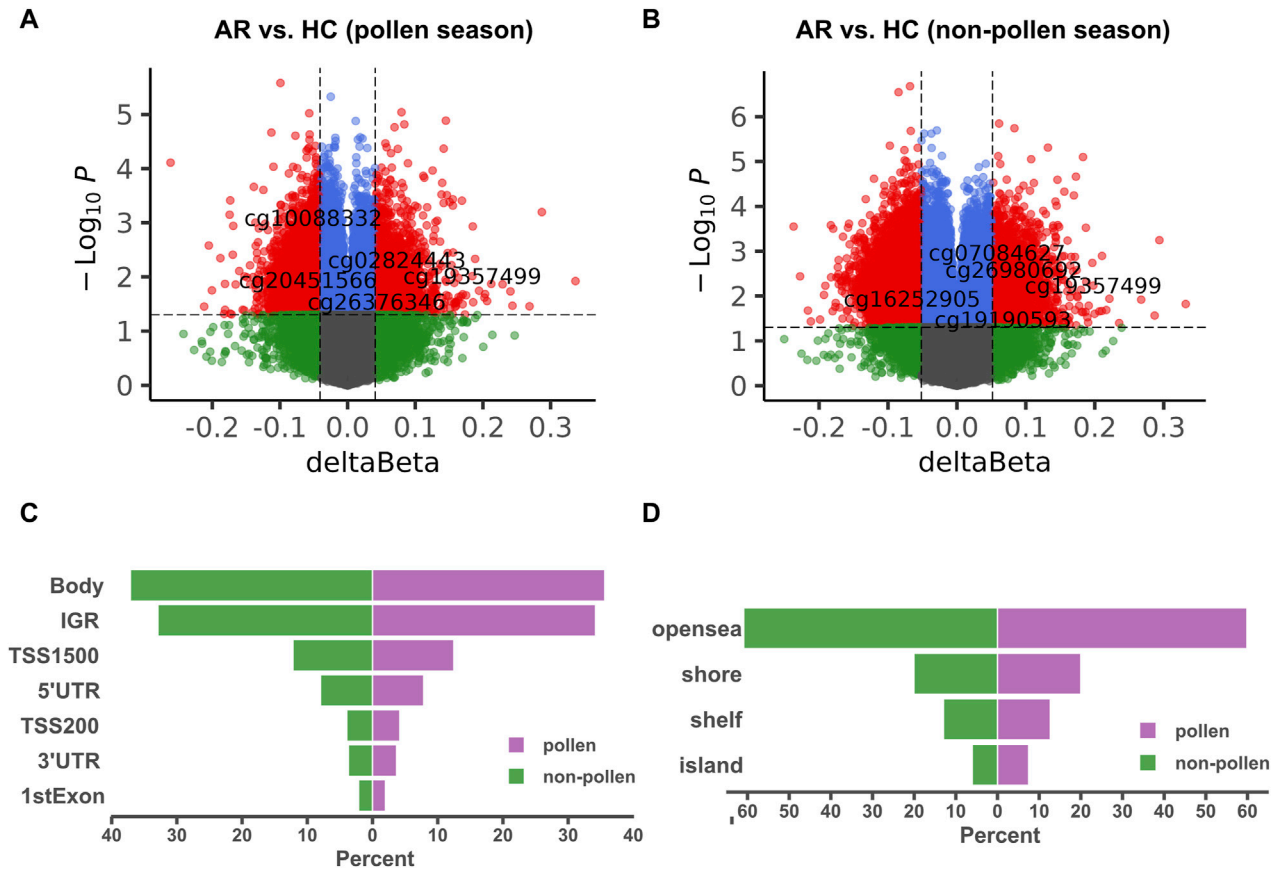


FIGURE 2
Identification of DMPs between AR and HC groups in GSE50222. **(A, B)** The volcano map of the DMPs during the pollen and non-pollen seasons. The X-axis corresponds to the β change value, while the Y-axis corresponds to the -$\log_{10}$(*p*-value). The red nodes show the hypomethylation and hypermethylation DMPs. **(C, D)** The distribution of the DMPs during the pollen and non-pollen seasons. The comparison was performed on exact promoter region and CpG island separately. DMP: differentially methylated position; AR: allergic rhinitis; HC: healthy control.

seasons when the filtering criterion was set as *p*-value < 0.05 and *Cor* < −0.4. By limiting CpG sites to DMPs and genes to DEGs, we identified 20 and 24 methylation-driven gene pairings during the pollen and non-pollen seasons. These CpG-Gene pairs include two types: hypermethylation low expression genes and hypomethylation high expression genes, as it is generally believed that methylation regulates transcription in a negative manner (Supplementary Tables S1, S2). The analysis revealed 11 unique genes associated with the pollen season and 16 unique genes associated with the non-pollen season, with four genes intersecting between them. The results of methylation-driven genes are shown in Figure 4A, B. The correlations of the four CpG-Gene pairs are shown in Figure 4C–F as examples. The expression level of these genes can be found in Supplementary Tables S3, S4.

Results of gene functional enrichment analysis

Supplementary Figure S3A–C depict the results of the gene function enrichment analysis of methylation-driven genes during the pollen season. Detailed information can be obtained in Supplementary Tables S5–S7. They were primarily enriched in the *Hippo* signaling pathway (hsa04390), Neutrophil degranulation (R-HSA-6798695), cell growth (GO:0016049), etc.

Supplementary Figure S3D–F show the results of the gene function enrichment analysis performed on methylation-driven genes during the non-pollen season. Detailed information can be obtained in Supplementary Tables S8–S10. They were primarily enriched in Cytokine-cytokine receptor interaction (hsa04060), Viral protein interaction with cytokine and cytokine receptor (hsa04061), Signaling by Interleukin (R-HSA-449147), cytokine-mediated signaling pathway (GO:0019221), etc.

Comparison results of methylation-driven genes with data on allergic comorbidity

The study analyzed the gene expression matrix of GSE101720, which includes 16,084 genes. Using this dataset, we investigated the DEGs in bronchial and nasal epithelial samples when comparing three groups: AR + asthma, AR, and HC. In bronchial epithelial samples, a total of 141 DEGs were identified, with 78 upregulated genes and 63 downregulated genes observed in the AR group compared to the HC group. When comparing the AR + asthma group to the HC

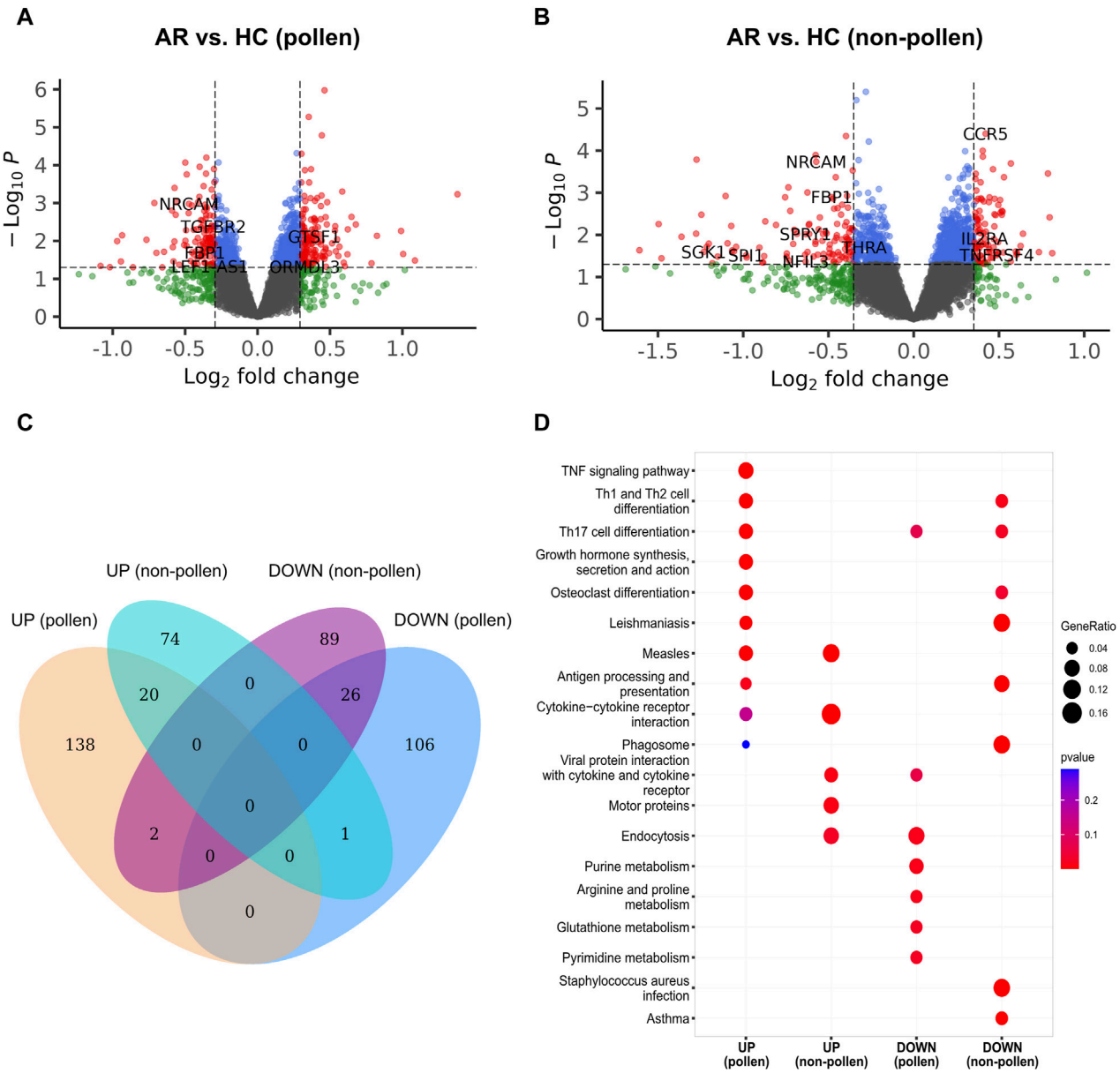


FIGURE 3
Identification of DEGs between AR and HC groups in GSE50101. **(A, B)** The volcano map of the DEGs during the pollen and non-pollen seasons. The X-axis corresponds to the $\log_2$FC, while the Y-axis corresponds to the -$\log_{10}$(*p*-value). The red node indicates the upregulated and downregulated DEGs. **(C)** The Venn plot of upregulated and downregulated DEGs during the pollen and non-pollen seasons. UP (pollen) and DOWN (pollen) represent the upregulated and downregulated DEGs of the pollen season. UP (non-pollen) and DOWN (non-pollen) represent the upregulated and downregulated DEGs of the non-pollen season. **(D)** KEGG pathways enriched by the subset of DEGs including UP (pollen), UP (non-pollen), DOWN (pollen), and DOWN (non-pollen). DEG: differentially expressed gene; AR: allergic rhinitis; HC: healthy control.

group, 421 DEGs were found in bronchial epithelial samples, including 220 upregulated genes and 201 downregulated genes. Similarly, in nasal epithelial samples, comparison of the AR group with the HC group revealed 313 DEGs, with 162 upregulated genes and 151 downregulated genes. Upon comparing the AR + asthma group with the HC group, we identified 314 DEGs consisting of 145 upregulated genes and 169 downregulated genes in nasal epithelial samples. Volcano plots that illustrate the DEGs analysis results are presented in Figure 5A–D.

Our previous results revealed that *ZNF667-AS1* was a methylation-driven gene unique to the pollen season. It was found downregulated in both AR and AR + asthma patients in nasal epithelial samples and downregulated in AR + asthma patients in bronchial epithelial samples. The inter-group expression levels of the *ZNF667-AS1* gene in GSE101720 are presented in Figure 5E, F. Additionally, *NRCAM* was found to be a methylation-driven gene shared by pollen and non-pollen periods. It was downregulated in both AR and AR + asthma patients in nasal epithelial samples.

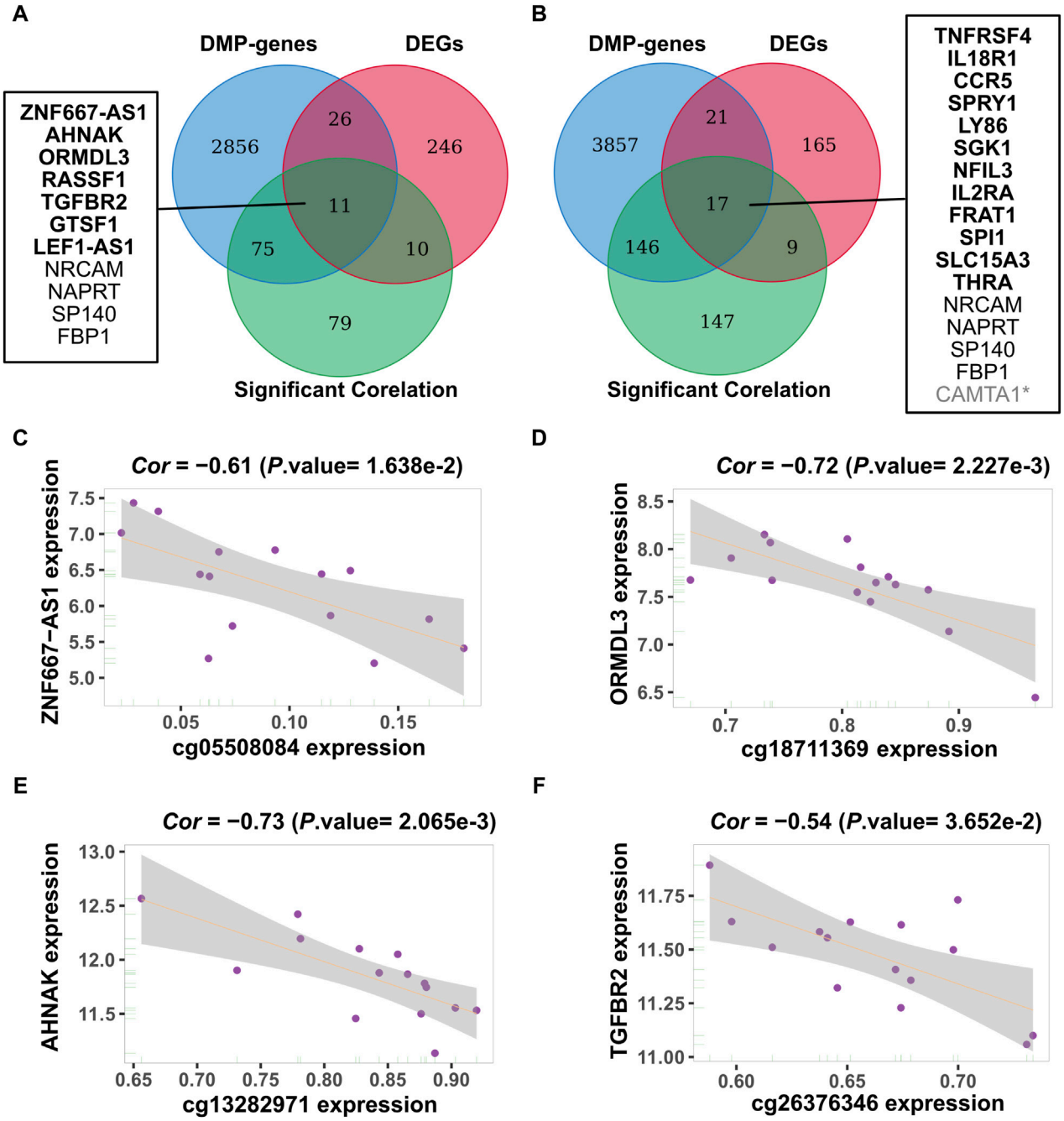


FIGURE 4
Identification of methylation-driven genes. **(A, B)** Identification of methylation-driven genes during the pollen and non-pollen seasons. The ellipses in three colors represent the DMP related gene set, the DEG gene set, and the gene set that exhibits a strong association between changes in methylation and expression. The middle part refers to the intersection of the three groups of data. The genes with the bolded name are specific methylation-driven genes. **(C–F)** Four specific methylation-driven genes of pollen season with correlation coefficient. The methylation-expression relationship was assessed using Spearman's correlation coefficient, with methylation and expression levels plotted on the X-axis and Y-axis, respectively. DMP: differentially methylated position; DEG: differentially expressed gene. * Due to the fact that the methylated probes negatively correlated with the expression of *CAMTA1* are not differentially expressed, *CAMTA1* was ultimately excluded from the list of methylation-driven genes.

Comparison results of methylation-driven genes with data on olive pollen allergy

The study analyzed the gene expression matrix of GSE54522, which includes 54,675 probes. Using this dataset, we investigated the DEGs in pollen-allergic patients and HC samples after olive pollen stimulation. A total of 1,321 DEGs were identified, with 1,206 upregulated genes and 115 downregulated genes observed in the allergic group compared to the HC group. We observed that four genes from the methylation-driven genes during the pollen season were also present among these DEGs. *ZNF667-AS1* was downregulated, while *AHNAK*, *ORMDL3*, and *RASSF1* were upregulated. Similarly, among the methylation-driven

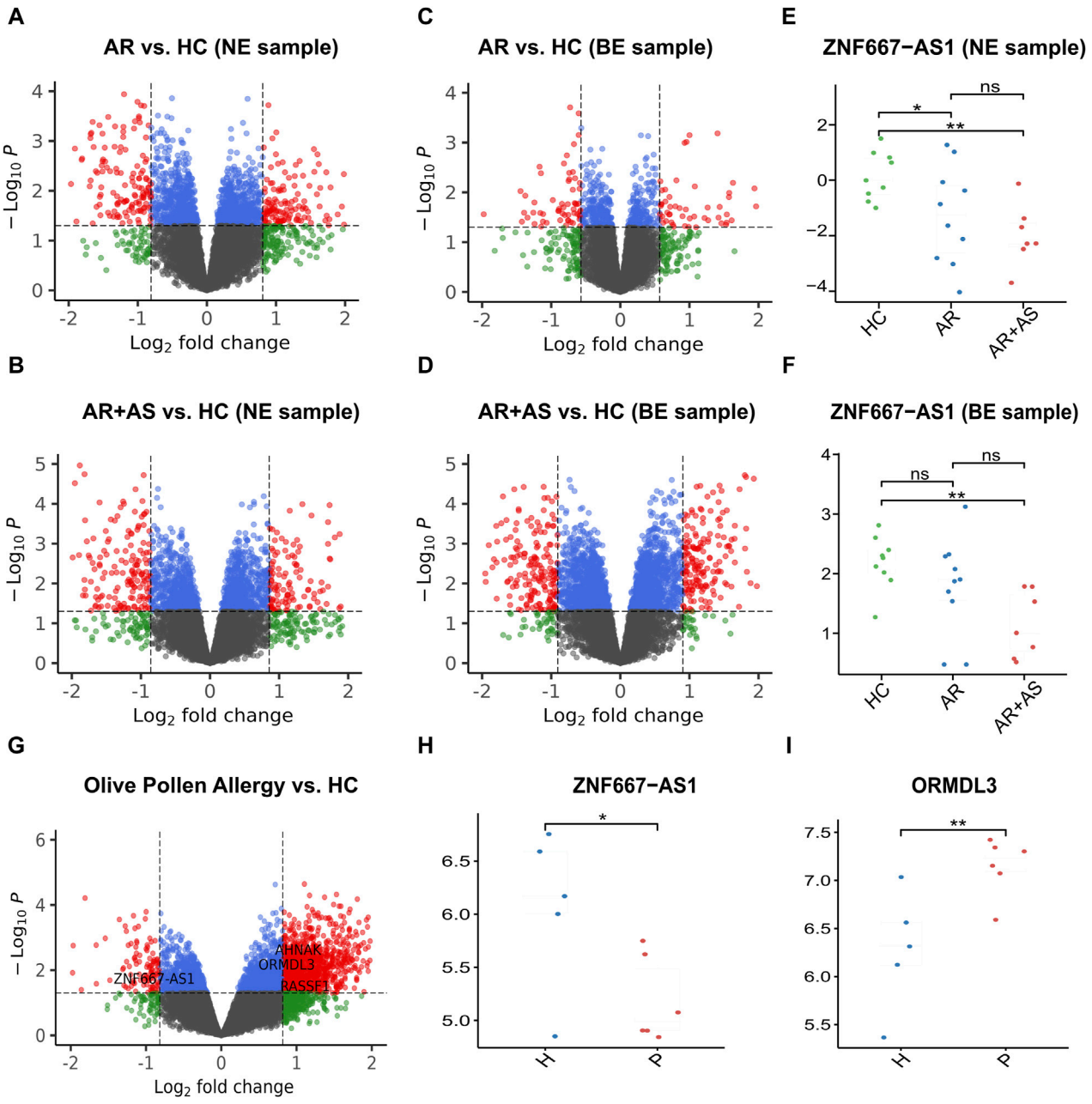


FIGURE 5
Result plot of validation using GSE101720 and GSE54522. **(A)** Volcano plot of AR vs. HC in GSE101720 (NE sample). **(B)** Volcano plot of AR with asthma vs. HC in GSE101720 (NE sample). **(C)** Volcano plot of AR vs. HC in GSE101720 (BE sample). **(D)** Volcano plot of AR with asthma vs. HC in GSE101720 (BE sample). **(E)** Expression level of *ZNF667-AS1* in GSE101720 (NE sample). **(F)** Expression level of *ZNF667-AS1* in GSE101720 (BE sample). **(G)** Volcano plot of olive pollen allergy patients vs. HC in GSE54522 (PBMC). **(H)** Expression level of *ZNF667-AS1* in GSE54522 (PBMC). **(I)** Expression level of *ORMDL3* in GSE54522 (PBMC). AR: allergic rhinitis; AS: asthma; HC: healthy control; NE: nasal epithelial; BE: bronchial epithelial.

genes during the non-pollen season, two genes, *SLC15A3* and *FRAT1*, were also identified as upregulated genes.

Results of weighted correlation network analysis for GSE50101

The pollen season expression matrix of GSE50101 contains a total of 47,314 genes and 19 samples. After probe annotation and removal of duplicates based on the highest expression value, 19,645 unique genes were remaining. Following filtering using the goodGenes function and Median Absolute Deviation, a total of 14,733 genes were included in the analysis. After filtering based on sample clustering and the goodSamples function, 18 samples were included in the analysis (Supplementary Figure S4A). The pickSoftThreshold function determined a soft threshold of 6, with R^2 cutoff of 0.85 (Figure 6A, B). The network construction resulted in 17 color modules (Figure 6C). Upon calculating the

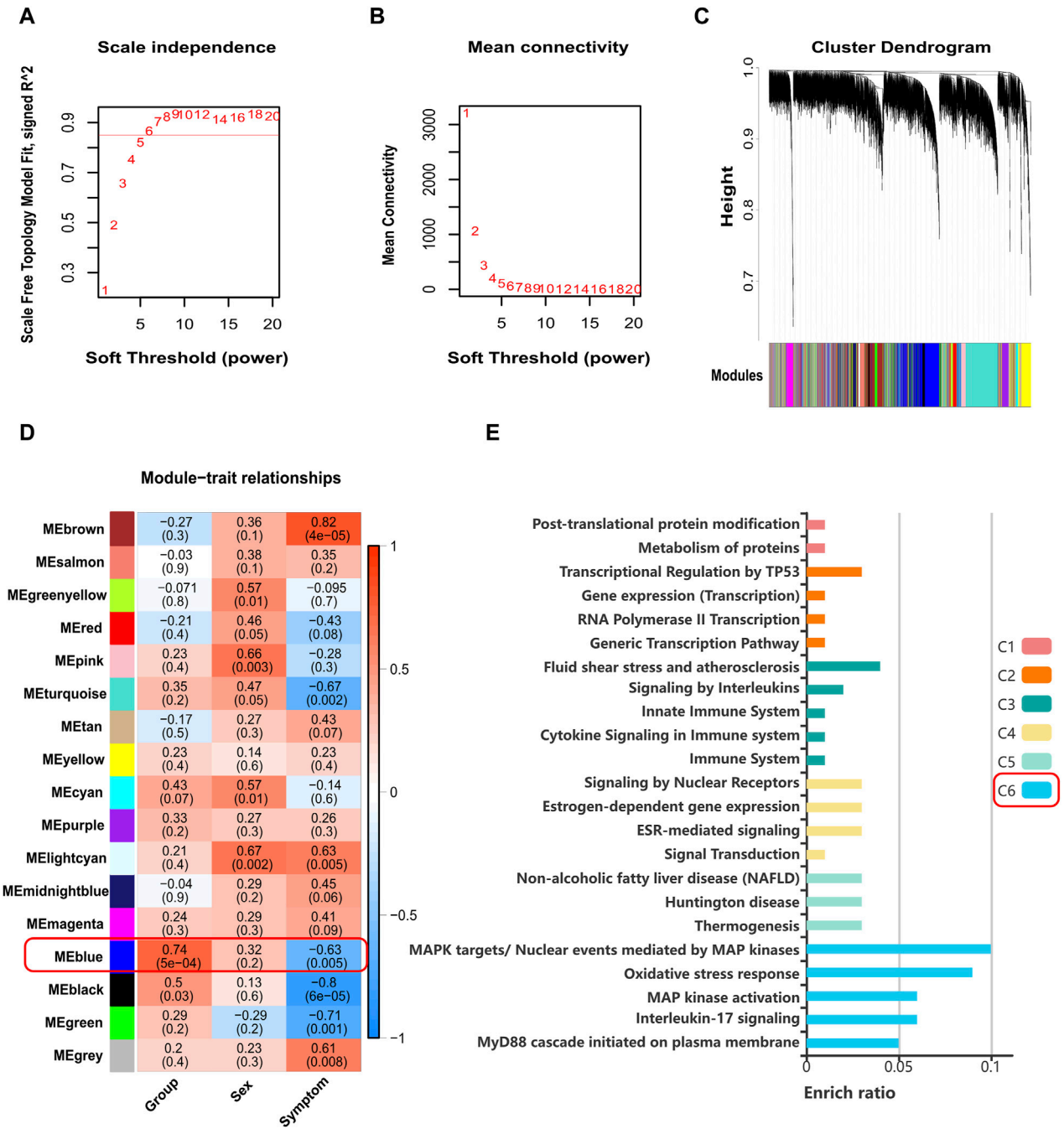


FIGURE 6
Results plot of WGCNA. **(A)** Scale independence and soft-threshold power. **(B)** Mean connectivity and soft-threshold power. **(C)** Cluster dendrogram of co-expressed genes. **(D)** Relationships of modules to traits. **(E)** Histogram of KOBAS-i pathway intelligent clustering for genes associated with *ZNF667-AS1* in the blue module. WGCNA: weighted correlation network analysis.

correlation between modules and traits, it was found that only the blue module showed significant correlation with grouping (*Cor* = 0.74, *p*-value < 0.05) (Figure 6D). The blue module consisted of 2,611 genes, and interestingly, *ZNF667-AS1* was also present in the blue module.

As WGCNA is a biological application used to identify highly correlated gene clusters (modules), and genes within these modules are presumed to exhibit coordinated interactions, we believe that the function of *ZNF667-AS1* can be investigated through studying the genes with which it exhibits co-expression within the blue module. The expression correlation between all genes in the blue module and *ZNF667-AS1* was calculated, resulting in 194 genes that showed significant correlation (*Cor* > 0.4, *p*-value < 0.05) (Supplementary Table S11). Pathway analysis using KOBAS-i was performed on these 195 genes (including *ZNF667-AS1*), and after filtering based on a Corrected *p*-value < 0.05, 7 pathways were enriched in KEGG, 41 pathways in Reactome, 1 pathway in

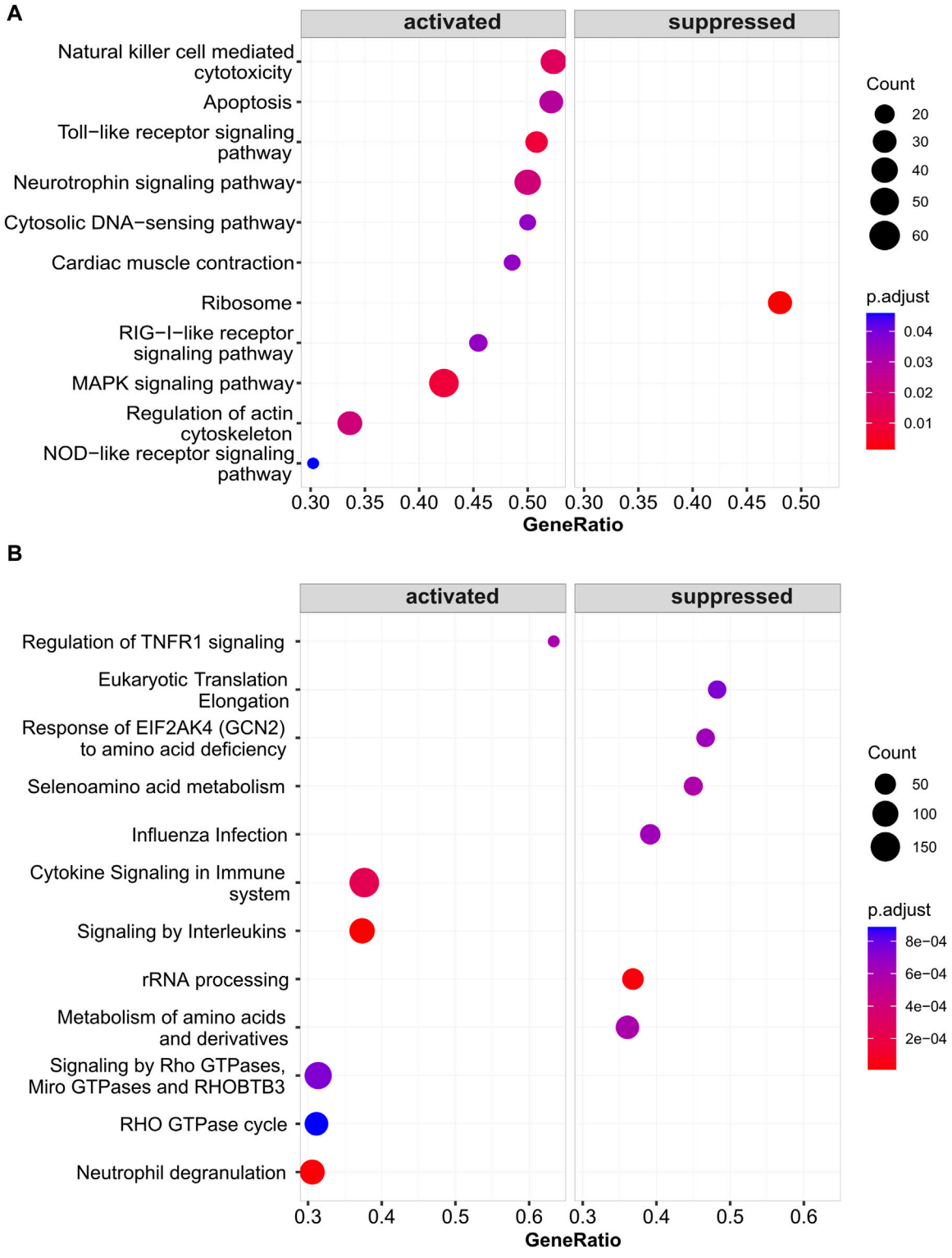


FIGURE 7
Results plot of Gene Set Enrichment Analysis for *cg05508084* (paired with *ZNF667-AS1*). **(A)** The KEGG pathways that are activated and inhibited by *cg05508084*. **(B)** The Reactome pathways that are activated and inhibited by *cg05508084*.

PANTHER, while no pathways were enriched in BioCyc. Intelligent clustering of the pathways revealed 6 main clusters, with Cluster C6 being the largest, consisting of 18 pathways. This cluster includes pathways such as Toll-like receptor (*TLR*) family, *Myd88*, *MAPK*, oxidative stress, and others, as depicted in Figure 6E and Supplementary Table S12.

Results of gene set enrichment analysis for *cg05508084*

The corresponding CpG site for *ZNF667-AS1* is *cg05508084*. GSEA analysis revealed that the functional significance of the *cg05508084* site may be associated with the activation of *TLR* receptor signaling pathway and *MAPK* signaling pathway. The results are shown in Figure 7A,B.

Discussion

Allergic rhinitis is a prevalent allergic airway disease characterized by symptoms such as sneezing, nasal congestion, nasal itching and rhinorrhea caused by immunoglobulin E (IgE)-mediated responses to inhaled allergens. These immune responses involve mucosal inflammation driven by type 2 cells, a subset of immune cells that produce cytokines responsible for inducing inflammation in response to allergens. Research suggests that exposure to allergens could trigger epigenetic changes (Cheng et al., 2014; Choi et al., 2021), leading to AR development with a susceptible genetic background (Bousquet et al., 2020a). Methylation has been shown to differentiate AR from the healthy individuals and to be associated with the development and exacerbation of AR (Li et al., 2016; Choi et al., 2021).

Pollen is one of the frequently occurring environmental allergens for AR (Skoner, 2001; Meng et al., 2020). Studies found that pollen-induced DNA methylation changes were correlated with symptom scores in AR patients, highlighting the role of epigenetic mechanisms in AR pathogenesis (Nestor et al., 2014; North et al., 2018; Watanabe et al., 2021; Tameeris et al., 2023). However, the specific effects of pollen season on AR patients have not been well elucidated. A bioinformatics study identified some specific signaling pathways for pollen season by comparing DMPs during the pollen and non-pollen seasons among AR patients and HCs (Yang et al., 2022). Our study goes further by integrating methylation profiling and gene expression profiling in a multi-omics approach, to identify methylation-driven genes specifically during the pollen and non-pollen seasons in AR to explore the unique epigenetic mechanisms of pollen season. This multi-omics approach of studying methylation-driven genes has been widely used in the study of various diseases, such as esophageal cancer (Jammula et al., 2020), breast cancer (Zhang et al., 2020), asthma (Chen et al., 2021; Wang et al., 2022), etc.

In our study, we identified 20 pairs of CpG-gene combinations in the pollen season and 24 such combinations in the non-pollen season. Due to the property of allergen-specific excitation in allergic diseases, we believe that the non-pollen season gene results and shared genes reflect the common AR methylation changes, while the results of the pollen season reflect pollen-specific AR methylation changes. According to the WGCNA results, the methylation-driven genes of pollen and non-pollen seasons were distributed in different color modules (Supplementary Tables S3, S4), the genes within the modules have synergistic effects, indicating that there might be distinct mechanism for methylation regulation in the pollen season.

We have identified 4 shared methylation-driven genes for the pollen and non-pollen seasons, namely, *SP140*, *NRCAM*, *NAPRT*, and *FBP1*. Their functions are associated with allergic diseases both on allergic inflammation and oxidative stress levels. *SP140* is critical for transcriptional programs that uphold the macrophage state and also is a potentially novel gene contributing to IgE-dependent mast cells (MCs) activation (Fraschilla et al., 2022). *Fbp1* can aggravate oxidative stress-induced apoptosis by suppressing *Nrf2* signaling, which exerts a significant impact on the prevalence and severity of asthma (Hu et al., 2021).

Methylation-driven genes specific to the non-pollen season include *TNFRSF4*, *IL18R1*, *CCR5*, *SPRY1*, *LY86*, *SGK1*, *NFIL3*, *IL2RA*, *FRAT1*, *SPI1*, *SLC15A3* and *THRA* (Figure 4B). These genes are primarily discussed at the pathway level, their functional enrichment results involve signaling pathways related to cytokines and interleukins, which include cytokine-cytokine receptor interaction (hsa04060), cytokine-mediated signaling pathway (GO:0019221), and signaling by interleukins (R-HSA-449147) (Supplementary Figure S3D–F). The pathways involved in this process regulate the complex signaling of allergic inflammation by influencing the recruitment and dissipation of cytokines and interleukins, as well as downstream reactions. Our team's previous research identified cytokine-cytokine receptor interaction as the key target pathway of Tuomin-Zhiti-Decoction for AR through proteomics and functional enrichment analysis subsequently (Cheng et al., 2022).

We concentrate on the methylation effects caused by pollen. Methylation-driven genes specific to the pollen season include *ZNF667-AS1*, *AHNAK*, *ORMDL3*, *RASSF1*, *TGFBR2*, *GTSF1* and *LEF1-AS1* (Figure 4A). Through literature retrieval, it has been confirmed that 4 out of the 7 unique genes are regulated by DNA methylation in their transcription (Gao et al., 2018; Walter et al., 2018; Wu et al., 2018; Ma et al., 2020), which may reflect the response of AR to pollen exposure. Based on data on olive pollen allergy (GSE54522), the first four of these genes (*ZNF667-AS1*, *AHNAK*, *ORMDL3* and *RASSF1*) were confirmed (Figure 5G). Previous studies have reported their association with allergic diseases, mainly involving the influence on genetic susceptibility and regulation of allergic inflammation. Research findings indicate that *AHNAK* triggers an inflammatory reaction through the activation of MCs (Song et al., 2023). Numerous genome-wide association studies have identified *ORMDL3* as a gene associated with asthma susceptibility (Moffatt et al., 2007; Galanter et al., 2008; James et al., 2019; Ntontsi et al., 2021). *ORMDL3* is a transmembrane protein found in the endoplasmic reticulum that regulates sphingolipid synthesis. The molecular mechanisms underlying *ORMDL3*'s pathologic functions in asthma are connected to its evolutionarily conserved role in the regulation of sphingolipid homeostasis (James et al., 2019). *TGFBR2* (Transforming Growth Factor Beta Receptor 2) is one of the major components of the transforming growth factor β (*TGFβ*) signaling (Weiss and Attisano, 2013). People with *TGF* receptor mutations are significantly more likely to develop allergic disorders (Frischmeyer-Guerrerio et al., 2013). At the pathway level, we have identified the *Hippo* signaling pathway (hsa04390) and the neutrophil degranulation-associated pathway (R-HSA-6798695). Recent research has proved that the *Hippo* signaling pathway can regulate the macrophage population size by mediating the microenvironment, which may be the basis for its association with AR (Zhou et al., 2022). Although neutrophils are not traditionally regarded as components of type 2 immunity, a

growing number of studies confirm that neutrophils may contribute to the initiation of type 2 immune responses (Wark et al., 2002; Kämpe et al., 2011; Kämpe et al., 2012).

Motivated by the well-known asthma-related gene *ORMDL3*, we decided to investigate if the pollen season's methylation effects could affect AR and contribute to the development of asthma. There is growing evidence showing a strong association between AR and asthma. Both conditions are epidemiologically and pathophysiologically related (Bergeron and Hamid, 2005; Compalati et al., 2010; Bousquet et al., 2019), and AR is considered an independent risk factor for the development of asthma (Compalati et al., 2010; Acevedo-Prado et al., 2022). The key disease targets for the progression of AR to asthma have yet to be elucidated. Therefore, we introduced external data (GSE101720) to explore DNA methylation as the underlying mechanism of the progression from AR to asthma. As seen in the above volcano plot (Figure 5A–D), with the disease proceeding from AR to asthma, there was a significant increase in the amount of DEGs in the airway mucosa, and the trend is reversed in the nasal mucosa, though the difference was not significant, indicating a transfer of the inflammatory site from the nasal mucosa to the airway mucosa.

We found the *ZNF667-AS1* gene was significantly downregulated within both the AR and AR + asthma groups in data GSE101720 (Figure 5E, F). Referring back to the modifications in the gene expression profile data GSE50101, we observed that *ZNF667-AS1* was also a downregulated expression DEG and that it also had the most significant change in the expression of the seven methylation-driven genes specific to the pollen season (Supplementary Table S3), leading us to conclude that it was the primary gene mediating the pollen effect. *ZNF667-AS1* is classified as a long-stranded non-coding RNA (lncRNA) with multifaceted activities. lncRNAs were proven to play an integral role in the pathogenesis of allergic disease by regulating the differentiation and apoptosis of hematopoietic stem cells, bone marrow cells, and the activation of monocytes, macrophages, and dendritic cells in immune regulation (Cheng et al., 2022). *ZNF667-AS1* has been the focus of recent research on inflammatory conditions and cancer with an emphasis on its function in the control of inflammatory pathways (Di Fiore et al., 2021; Zheng et al., 2021; Fan et al., 2022; Luan et al., 2022; Ma et al., 2022; Bohosova et al., 2023). Excitingly, we discovered that Liu et al.'s bioinformatics study (Liu et al., 2019), based on data GSE67472, identified *ZNF667-AS1* as an important gene involved in the pathogenesis of asthma, which is also downregulated in asthma. We applied WGCNA to investigate the function of *ZNF667-AS1* further, as the functional annotation of lncRNA is still in its exploratory stages. The blue module was the only significant module related to AR (Cor = 0.74, p-value < 0.05) (Figure 6D) and *ZNF667-AS1* was the only methylation-driven gene in the blue module. Excited by this outcome, we proceeded to perform pathway enrichment analysis on the genes associated with *ZNF667-AS1* in the blue module. The reason for doing so is that genes identified within the same module by WGCNA are assumed to have high synergistic changes. This characteristic has been frequently used in recent years for functional annotation of lncRNAs, specifically exploring the function of lncRNAs through annotating the functions of mRNAs co-expressed with them (Wang et al., 2023). What we discovered was that the most significant class of KOBAS-i pathway intelligent clustering results comprised 18 pathways, including *TLR* family, *Myd88*, *MAPK*, and oxidative stress. *ZNF667-AS1* may also be implicated in activating *TLR* and *MAPK* signaling pathways, according to gene set enrichment study conducted around the CpG site *cg05508084* corresponding to *ZNF667-AS1*. Pollen can directly interact with pattern recognition receptors (PRRs), particularly *TLR* family, as many studies have proved to date. This interaction can induce *TSLP* production and related type 2 inflammation, and it may also be related to reactive oxygen-mediated oxidative stress (Hosoki et al., 2015; Pointner et al., 2020; Pointner et al., 2021; Shumin et al., 2021; Guryanova et al., 2022). Based on our research, methylation of the CpG site *cg05508084* in response to pollen stimulation drives the low expression of *ZNF667-AS1*, which in turn triggers type 2 inflammation and oxidative stress by affecting the *Myd88*-dependent *TLR* family, which may ultimately promote the development of AR to asthma. *ZNF667-AS1* may serve as a crucial mediator and biomarker of allergic diseases.

As per the concept of the "unified airway", the upper and lower respiratory tracts share common barrier and immune characteristics, making them a morphologically and functionally unified entity (Bachert et al., 2023; Wang et al., 2023), which can explain that allergy is not a disease of a specific organ, but a disorder of the whole respiratory tract (Compalati et al., 2010). This coincides with the concept of the allergic constitution, which is considered as the background and foundation for the development of allergic diseases by the theory of traditional Chinese medicine constitution (Wang et al., 2023). Our previous study (Sun et al., 2024) suggests that pathways such as T_H1 and T_H2 cell differentiation, *TLR* cascade are the key molecular characterization of allergic constitution. We believe that the occurrence of allergic disease is the result of a combination of innate constitution and environmental factors.

Dutch children and adolescents' overall symptom scores are positively correlated with grass and birch pollen concentrations, according to a recent study (Tameeris et al., 2023). We hypothesize that methylation effects could be associated with symptoms throughout the pollen season. According to the WGCNA results, blue, brown, black, green, turquoise, and lightcyan modules show a strong correlation with the allergy symptom score (Figure 6D). Moreover, a number of researches have documented a connection between pollen and allergy symptoms (Tobías et al., 2003; Kiotseridis et al., 2013; Kmenta et al., 2016; Wang et al., 2018). Unfortunately, because we do not have pollen concentration information in our data sources, we cannot analyze whether it correlates with methylation levels. That is one of our study's limitations. Our analysis method is mainly based on bioinformatics, the dependability of the results is obviously affected by the sample size and the depth of the sample data. Furthermore, the data we used was restricted to transcriptomics and epigenomics, further research is required at multiple omics levels, including proteomics and metabolomics. In order to uncover the long-term effects of pollen season on AR at a deeper level, it is intended that additional research can be conducted in the future.

Conclusion

In this study, we identified the methylation-driven genes triggered by pollen and investigated their functions through bioinformatic analysis. Four of the seven key methylation-driven genes (*ZNF667-AS1*, *AHNAK*, *ORMDL3* and *RASSF1*) were

validated by another olive pollen allergy data. By incorporating external data, we identified *ZNF667-AS1* as a crucial methylation-driven gene mediating pollen effects, which could be a key gene driving the conversion of AR to asthma. Its potential mechanism may involve the influence on *TLR* family, further mediating type 2 inflammatory responses and impacting the occurrence and development of AR and asthma. Our findings provide new insights into the interaction between AR and asthma, and pave the way for scientific research and development of therapeutic strategies. We hope there will be researches conducted *in vitro* and *in vivo* in the future to validate our findings.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

Author contributions

PS and YW conceived and designed the study. PS and XL performed the data analysis and data interpretation. YW and ZL discussed the results and provided additional information. DC and QL plotted the figures. YW and PS prepared the manuscript. JW and QW revised and finalized the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the General project of Beijing Natural Science Foundation (No. 7242227), the Fundamental Research Funds for the Central Universities (No.2023-JYB-JBZD-009), the High-level Key Discipline of National Administration of Traditional Chinese Medicine–Traditional Chinese constitutional medicine (No.zyyzdxk-2023251), and the General program of National Natural Science Foundation of China (No. 82174243).

Acknowledgments

We extend our gratitude to the GEO database for offering their platform and to all the contributors who have uploaded their valuable datasets.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2024.1242974/full#supplementary-material

References

Acevedo-Prado, A., Seoane-Pillado, T., López-Silvarrey-Varela, A., Salgado, F.-J., Cruz, M.-J., Faraldo-Garcia, A., et al. (2022). Association of rhinitis with asthma prevalence and severity. *Sci. Rep.* 12, 6389. doi:10.1038/s41598-022-10448-w

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29. doi:10.1038/75556

Bachert, C., Luong, A. U., Gevaert, P., Mullol, J., Smith, S. G., Silver, J., et al. (2023). The unified airway hypothesis: evidence from specific intervention with anti-IL-5 biologic therapy. *J. Allergy Clin. Immunol. Pract.* S2213-2198 (23), 2630–2641. doi:10.1016/j.jaip.2023.05.011

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* 41, D991–D995. doi:10.1093/nar/gks1193

Bergeron, C., and Hamid, Q. (2005). Relationship between asthma and rhinitis: epidemiologic, pathophysiologic, and therapeutic aspects. *Allergy Asthma Clin. Immunol.* 1, 81–87. doi:10.1186/1710-1492-1-2-81

Bohosova, J., Kozelkova, K., Al Tukmachi, D., Trachtova, K., Naar, O., Ruckova, M., et al. (2023). Long non-coding RNAs enable precise diagnosis and prediction of early relapse after nephrectomy in patients with renal cell carcinoma. *J. Cancer Res. Clin. Oncol.* 149, 7587–7600. doi:10.1007/s00432-023-04700-7

Bousquet, J., Anto, J. M., Bachert, C., Baiardini, I., Bosnic-Anticevich, S., Walter Canonica, G., et al. (2020a). Allergic rhinitis. *Nat. Rev. Dis. Prim.* 6, 95–17. doi:10.1038/s41572-020-00227-0

Bousquet, J., Hellings, P. W., Agache, I., Amat, F., Annesi-Maesano, I., Ansotegui, I. J., et al. (2019). Allergic Rhinitis and its Impact on Asthma (ARIA) Phase 4 (2018): change management in allergic rhinitis and asthma multimorbidity using mobile technology. *J. Allergy Clin. Immunol.* 143, 864–879. doi:10.1016/j.jaci.2018.08.049

Bousquet, J., Schünemann, H. J., Togias, A., Bachert, C., Erhola, M., Hellings, P. W., et al. (2020b). Next-generation allergic rhinitis and its impact on asthma (ARIA) guidelines for allergic rhinitis based on grading of recommendations assessment, development and evaluation (GRADE) and real-world evidence. *J. Allergy Clin. Immunol.* 145, 70–80.e3. doi:10.1016/j.jaci.2019.06.049

Bu, D., Luo, H., Huo, P., Wang, Z., Zhang, S., He, Z., et al. (2021). KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res.* 49, W317–W325. doi:10.1093/nar/gkab447

Calzada, D., Aguerri, M., Baos, S., Montaner, D., Mata, M., Dopazo, J., et al. (2015). Therapeutic targets for olive pollen allergy defined by gene markers modulated by Ole e 1-derived peptides. *Mol. Immunol.* 64, 252–261. doi:10.1016/j.molimm.2014.12.002

Carlson, M. (2021). org.Hs.eg.db: genome wide annotation for Human. Available at: https://bioconductor.org/packages/org.Hs.eg.db/.

Chen, H. (2018). VennDiagram: generate high-resolution venn and euler plots. Available at: https://CRAN.R-project.org/package=VennDiagram.

Chen, R., Piao, L.-Z., Liu, L., and Zhang, X.-F. (2021). DNA methylation and gene expression profiles to identify childhood atopic asthma associated genes. *BMC Pulm. Med.* 21, 292. doi:10.1186/s12890-021-01655-8

Chen, Y., Lemire, M., Choufani, S., Butcher, D. T., Grafodatskaya, D., Zanke, B. W., et al. (2013). Discovery of cross-reactive probes and polymorphic CpGs in the Illumina infinium HumanMethylation450 microarray. *Epigenetics* 8, 203–209. doi:10.4161/epi.23470

Cheng, J., Zhang, M., Zheng, Y., Wang, J., and Wang, Q. (2022a). Integrative analysis of network pharmacology and proteomics to identify key targets of Tuomin-Zhiti-Decoction for allergic rhinitis. *J. Ethnopharmacol.* 296, 115448. doi:10.1016/j.jep.2022.115448

Cheng, R. Y., Shang, Y., Limjunyawong, N., Dao, T., Das, S., Rabold, R., et al. (2014). Alterations of the lung methylome in allergic airway hyper-responsiveness. *Environ. Mol. Mutagen* 55, 244–255. doi:10.1002/em.21851

Cheng, S., Tang, Q., Xie, S., Wen, S., Zhang, H., Xie, Z., et al. (2022b). The role of noncoding RNA in airway allergic diseases through regulation of T cell subsets. *Mediat. Inflamm.* 2022, 6125698. doi:10.1155/2022/6125698

Choi, B. Y., Han, M., Kwak, J. W., and Kim, T. H. (2021). Genetics and epigenetics in allergic rhinitis. *Genes (Basel)* 12, 2004. doi:10.3390/genes12122004

Compalati, E., Ridolo, E., Passalacqua, G., Braido, F., Villa, E., and Canonica, G. W. (2010). The link between allergic rhinitis and asthma: the united airways disease. *Expert Rev. Clin. Immunol.* 6, 413–423. doi:10.1586/eci.10.15

Davis, S., and Meltzer, P. S. (2007). GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor. *Bioinformatics* 23, 1846–1847. doi:10.1093/bioinformatics/btm254

Dharmage, S. C., Lowe, A. J., and Tang, M. L. K. (2022). Revisiting the atopic march current evidence. *Am. J. Respir. Crit. Care Med.* 206, 925–926. doi:10.1164/rccm.202206-1219ED

Di Fiore, R., Suleiman, S., Drago-Ferrante, R., Felix, A., O'Toole, S. A., O'Leary, J. J., et al. (2021). LncRNA MORT (ZNF667-AS1) in cancer-is there a possible role in gynecological malignancies? *Int. J. Mol. Sci.* 22, 7829. doi:10.3390/ijms22157829

Dolgalev, I. (2022). Msigdbr: MSigDB gene sets for multiple organisms in a tidy data format. Available at: https://CRAN.R-project.org/package=msigdbr.

Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30, 207–210. doi:10.1093/nar/30.1.207

Fan, X., Lu, P., Wang, H., Bian, S., Wu, X., Zhang, Y., et al. (2022). Integrated single-cell multiomics analysis reveals novel candidate markers for prognosis in human pancreatic ductal adenocarcinoma. *Cell Discov.* 8, 13. doi:10.1038/s41421-021-00366-y

Fraschilla, I., Amatullah, H., Rahman, R.-U., and Jeffrey, K. L. (2022). Immune chromatin reader SP140 regulates microbiota and risk for inflammatory bowel disease. *Cell Host Microbe* 30, 1370–1381.e5. doi:10.1016/j.chom.2022.08.018

Frischmeyer-Guerrerio, P. A., Guerrerio, A. L., Oswald, G., Chichester, K., Myers, L., Halushka, M. K., et al. (2013). TGFβ receptor mutations impose a strong predisposition for human allergic disease. *Sci. Transl. Med.* 5, 195ra94. doi:10.1126/scitranslmed.3006448

Galanter, J., Choudhry, S., Eng, C., Nazario, S., Rodríguez-Santana, J. R., Casal, J., et al. (2008). *ORMDL3* gene is associated with asthma in three ethnically diverse populations. *Am. J. Respir. Crit. Care Med.* 177, 1194–1200. doi:10.1164/rccm.200711-1644OC

Gao, D.-Y., Ling, Y., Lou, X.-L., Wang, Y.-Y., and Liu, L.-M. (2018). GTSF1 gene may serve as a novel potential diagnostic biomarker for liver cancer. *Oncol. Lett.* 15, 3133–3140. doi:10.3892/ol.2017.7695

García-Mozo, H. (2017). Poaceae pollen as the leading aeroallergen worldwide: a review. *Allergy* 72, 1849–1858. doi:10.1111/all.13210

Gene Ontology Consortium (2021). The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.* 49, D325–D334. doi:10.1093/nar/gkaa1113

Gillespie, M., Jassal, B., Stephan, R., Milacic, M., Rothfels, K., Senff-Ribeiro, A., et al. (2022). The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* 50, D687–D692. doi:10.1093/nar/gkab1028

Giovannini-Chami, L., Paquet, A., Sanfiorenzo, C., Pons, N., Cazareth, J., Magnone, V., et al. (2018). The "one airway, one disease" concept in light of Th2 inflammation. *Eur. Respir. J.* 52, 1800437. doi:10.1183/13993003.00437-2018

Greiner, A. N., Hellings, P. W., Rotiroti, G., and Scadding, G. K. (2011). Allergic rhinitis. *Lancet* 378, 2112–2122. doi:10.1016/S0140-6736(11)60130-X

Guo, Y., Mao, X., Qiao, Z., Chen, B., and Jin, F. (2020). A novel promoter CpG-based signature for long-term survival prediction of breast cancer patients. *Front. Oncol.* 10, 579692. doi:10.3389/fonc.2020.579692

Guryanova, S. V., Finkina, E. I., Melnikova, D. N., Bogdanov, I. V., Bohle, B., and Ovchinnikova, T. V. (2022). How do pollen allergens sensitize? *Front. Mol. Biosci.* 9, 900533. doi:10.3389/fmolb.2022.900533

Hosoki, K., Boldogh, I., and Sur, S. (2015). Innate responses to pollen allergens. *Curr. Opin. Allergy Clin. Immunol.* 15, 79–88. doi:10.1097/ACI.0000000000000136

Hu, J., Wang, J., Li, C., and Shang, Y. (2021). Fructose-1,6-bisphosphatase aggravates oxidative stress-induced apoptosis in asthma by suppressing the Nrf2 pathway. *J. Cell Mol. Med.* 25, 5001–5014. doi:10.1111/jcmm.16439

James, B., Milstien, S., and Spiegel, S. (2019). ORMDL3 and allergic asthma: from physiology to pathology. *J. Allergy Clin. Immunol.* 144, 634–640. doi:10.1016/j.jaci.2019.07.023

Jammula, S., Katz-Summercorn, A. C., Li, X., Linossi, C., Smyth, E., Killcoyne, S., et al. (2020). Identification of subtypes of barrett's esophagus and esophageal adenocarcinoma based on DNA methylation profiles and integration of transcriptome and genome data. *Gastroenterology* 158, 1682–1697.e1. doi:10.1053/j.gastro.2020.01.044

Kämpe, M., Lampinen, M., Stolt, I., Janson, C., Stålenheim, G., and Carlson, M. (2012). PI3-kinase regulates eosinophil and neutrophil degranulation in patients with allergic rhinitis and allergic asthma irrespective of allergen challenge model. *Inflammation* 35, 230–239. doi:10.1007/s10753-011-9309-5

Kämpe, M., Stolt, I., Lampinen, M., Janson, C., Stålenheim, G., and Carlson, M. (2011). Patients with allergic rhinitis and allergic asthma share the same pattern of eosinophil and neutrophil degranulation after allergen challenge. *Clin. Mol. Allergy* 9, 3. doi:10.1186/1476-7961-9-3

Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2021). KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* 49, D545–D551. doi:10.1093/nar/gkaa970

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi:10.1093/nar/28.1.27

Kiotseridis, H., Cilio, C. M., Bjermer, L., Tunsäter, A., Jacobsson, H., and Dahl, A. (2013). Grass pollen allergy in children and adolescents-symptoms, health related quality of life and the value of pollen prognosis. *Clin. Transl. Allergy* 3, 19. doi:10.1186/2045-7022-3-19

Kmenta, M., Bastl, K., Kramer, M. F., Hewings, S. J., Mwange, J., Zetter, R., et al. (2016). The grass pollen season 2014 in Vienna: A pilot study combining phenology, aerobiology and symptom data. *Sci. Total Environ.* 566 (567), 1614–1620. doi:10.1016/j.scitotenv.2016.06.059

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

Law, P.-P., and Holland, M. L. (2019). DNA methylation at the crossroads of gene and environment interactions. *Essays Biochem.* 63, 717–726. doi:10.1042/EBC20190031

Li, J. Y., Zhang, Y., Lin, X. P., Ruan, Y., Wang, Y., Wang, C. S., et al. (2016). Association between DNA hypomethylation at IL13 gene and allergic rhinitis in house dust mite-sensitized subjects. *Clin. Exp. Allergy* 46, 298–307. doi:10.1111/cea.12647

Liu, X., Zhang, Y., Jiang, H., Jiang, N., and Gao, J. (2019). Integrative analysis of the contribution of mRNAs and long non-coding RNAs to the pathogenesis of asthma. *Mol. Med. Rep.* 20, 2617–2624. doi:10.3892/mmr.2019.10511

Luan, L., Dai, Y., Shen, T., Yang, C., Chen, Z., Liu, S., et al. (2022). Development of a novel hypoxia-immune-related LncRNA risk signature for predicting the prognosis and immunotherapy response of colorectal cancer. *Front. Immunol.* 13, 951455. doi:10.3389/fimmu.2022.951455

Luo, Y., Zeng, G., and Wu, S. (2019). Identification of microenvironment-related prognostic genes in bladder cancer based on gene expression profile. *Front. Genet.* 10, 1187. doi:10.3389/fgene.2019.01187

Ma, X., Yu, S., Zhao, B., Bai, W., Cui, Y., Ni, J., et al. (2022). Development and validation of a novel ferroptosis-related LncRNA signature for predicting prognosis and the immune landscape features in uveal melanoma. *Front. Immunol.* 13, 922315. doi:10.3389/fimmu.2022.922315

Ma, Y., He, S., Gao, A., Zhang, Y., Zhu, Q., Wang, P., et al. (2020). Methylation silencing of TGF-β receptor type II is involved in malignant transformation of esophageal squamous cell carcinoma. *Clin. Epigenetics* 12, 25. doi:10.1186/s13148-020-0819-6

Meng, Y., Wang, C., and Zhang, L. (2020). Advances and novel developments in allergic rhinitis. *Allergy* 75, 3069–3076. doi:10.1111/all.14586

Miller, R. L., Grayson, M. H., and Strothman, K. (2021). Advances in asthma: new understandings of asthma's natural history, risk factors, underlying mechanisms, and clinical management. *J. Allergy Clin. Immunol.* 148, 1430–1441. doi:10.1016/j.jaci.2021.10.001

Moffatt, M. F., Kabesch, M., Liang, L., Dixon, A. L., Strachan, D., Heath, S., et al. (2007). Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 448, 470–473. doi:10.1038/nature06014

Morris, T. J., Butcher, L. M., Feber, A., Teschendorff, A. E., Chakravarthy, A. R., Wojdacz, T. K., et al. (2014). ChAMP: 450k chip analysis methylation pipeline. *Bioinformatics* 30, 428–430. doi:10.1093/bioinformatics/btt684

Nappi, E., Paoletti, G., Malvezzi, L., Ferri, S., Racca, F., Messina, M. R., et al. (2022). Comorbid allergic rhinitis and asthma: important clinical considerations. *Expert Rev. Clin. Immunol.* 18, 747–758. doi:10.1080/1744666X.2022.2089654

Nestor, C. E., Barrenäs, F., Wang, H., Lentini, A., Zhang, H., Bruhn, S., et al. (2014). DNA methylation changes separate allergic patients from healthy controls and may reflect altered CD4+ T-cell population structure. *PLoS Genet.* 10, e1004059. doi:10.1371/journal.pgen.1004059

North, M. L., Jones, M. J., MacIsaac, J. L., Morin, A. M., Steacy, L. M., Gregor, A., et al. (2018). Blood and nasal epigenetics correlate with allergic rhinitis symptom development in the environmental exposure unit. *Allergy* 73, 196–205. doi:10.1111/all.13263

Ntontsi, P., Photiades, A., Zervas, E., Xanthou, G., and Samitas, K. (2021). Genetics and epigenetics in asthma. *IJMS* 22, 2412. doi:10.3390/ijms22052412

Paller, A. S., Spergel, J. M., Mina-Osorio, P., and Irvine, A. D. (2019). The atopic march and atopic multimorbidity: many trajectories, many pathways. *J. Allergy Clin. Immunol.* 143, 46–55. doi:10.1016/j.jaci.2018.11.006

Pidsley, R., Zotenko, E., Peters, T. J., Lawrence, M. G., Risbridger, G. P., Molloy, P., et al. (2016). Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol.* 17, 208. doi:10.1186/s13059-016-1066-1

Pointner, L., Bethanis, A., Thaler, M., Traidl-Hoffmann, C., Gilles, S., Ferreira, F., et al. (2020). Initiating pollen sensitization – complex source, complex mechanisms. *Clin. Transl. Allergy* 10, 36. doi:10.1186/s13601-020-00341-y

Pointner, L., Kraiem, A., Thaler, M., Richter, F., Wenger, M., Bethanis, A., et al. (2021). Birch pollen induces toll-like receptor 4-dependent dendritic cell activation favoring T cell responses. *Front. Allergy* 2, 680937. doi:10.3389/falgy.2021.680937

Qi, C., Jiang, Y., Yang, I. V., Forno, E., Wang, T., Vonk, J. M., et al. (2020). Nasal DNA methylation profiling of asthma and rhinitis. *J. Allergy Clin. Immunol.* 145, 1655–1663. doi:10.1016/j.jaci.2019.12.911

R Core Team (2023). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

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, e47. doi:10.1093/nar/gkv007

Schutzmeier, P., Kutzora, S., Mittermeier, I., Becker, J., Bergmann, K., Böse-O'Reilly, S., et al. (2022). Non-pharmacological interventions for pollen-induced allergic symptoms: systematic literature review. *Pediatr. Allergy Immunol.* 33, e13690. doi:10.1111/pai.13690

Shumin, Z., Luying, Z., Senlin, L., Jiaxian, P., Yang, L., Lanfang, R., et al. (2021). Ambient particulate matter-associated autophagy alleviates pulmonary inflammation induced by Platanus pollen protein 3 (Pla3). *Sci. Total Environ.* 758, 143696. doi:10.1016/j.scitotenv.2020.143696

Skoner, D. P. (2001). Allergic rhinitis: definition, epidemiology, pathophysiology, detection, and diagnosis. *J. Allergy Clin. Immunol.* 108, S2–S8. doi:10.1067/mai.2001.115569

Song, X., Zhang, L., Du, X., Zheng, Y., Jia, T., Zhou, T., et al. (2023). Neuroblast differentiation-associated protein derived polypeptides: AHNAK(5758-5775) induces inflammation by activating mast cells via ST2. *Immunol. Invest.* 52, 178–193. doi:10.1080/08820139.2022.2151368

Ständer, S. (2021). Atopic dermatitis. *N. Engl. J. Med.* 384, 1136–1143. doi:10.1056/NEJMra2023911

Suanno, C., Aloisi, I., Fernández-González, D., and Del Duca, S. (2021). Pollen forecasting and its relevance in pollen allergen avoidance. *Environ. Res.* 200, 111150. doi:10.1016/j.envres.2021.111150

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* 102, 15545–15550. doi:10.1073/pnas.0506580102

Sun, P., Liu, X., Wang, Y., Shen, R., Chen, X., Li, Z., et al. (2024). Molecular characterization of allergic constitution based on network pharmacology and multi-omics analysis methods. *Medicine* 103, e36892. doi:10.1097/MD.0000000000036892

Tameeris, E., Bohnen, A. M., Bindels, P. J. E., and Elshout, G. (2023). Clinical response to varying pollen exposure in allergic rhinitis in children in The Netherlands. *BMC Pediatr.* 23, 258. doi:10.1186/s12887-023-04021-1

Taylor, P. E., Jacobson, K. W., House, J. M., and Glovsky, M. M. (2007). Links between pollen, atopy and the asthma epidemic. *Int. Arch. Allergy Immunol.* 144, 162–170. doi:10.1159/000103230

Tian, Y., Morris, T. J., Webster, A. P., Yang, Z., Beck, S., Feber, A., et al. (2017). ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics* 33, 3982–3984. doi:10.1093/bioinformatics/btx513

Tobías, A., Galán, I., Banegas, J. R., and Aránguez, E. (2003). Short term effects of airborne pollen concentrations on asthma epidemic. *Thorax* 58, 708–710. doi:10.1136/thorax.58.8.708

Walter, R. F. H., Rozynek, P., Casjens, S., Werner, R., Mairinger, F. D., Speel, E. J. M., et al. (2018). Methylation of L1RE1, RARB, and RASSF1 function as possible biomarkers for the differential diagnosis of lung cancer. *PLoS One* 13, e0195716. doi:10.1371/journal.pone.0195716

Wang, J., Zhou, Y., Zhang, H., Hu, L., Liu, J., Wang, L., et al. (2023a). Pathogenesis of allergic diseases and implications for therapeutic interventions. *Signal Transduct. Target Ther.* 8, 138. doi:10.1038/s41392-023-01344-4

Wang, M., Gong, L., Luo, Y., He, S., Zhang, X., Xie, X., et al. (2023b). Transcriptomic analysis of asthma and allergic rhinitis reveals CST1 as a biomarker of unified airways. *Front. Immunol.* 14, 1048195. doi:10.3389/fimmu.2023.1048195

Wang, P., Paquet, É. R., and Robert, C. (2023c). Comprehensive transcriptomic analysis of long non-coding RNAs in bovine ovarian follicles and early embryos. *PLoS ONE* 18, e0291761. doi:10.1371/journal.pone.0291761

Wang, X.-Y., Ma, T.-T., Wang, X.-Y., Zhuang, Y., Wang, X.-D., Ning, H.-Y., et al. (2018). Prevalence of pollen-induced allergic rhinitis with high pollen exposure in grasslands of northern China. *Allergy* 73, 1232–1243. doi:10.1111/all.13388

Wang, Z., Wang, L., Dai, L., Wang, Y., Li, E., An, S., et al. (2022). Identification of candidate aberrant differentially methylated/expressed genes in asthma. *Allergy Asthma Clin. Immunol.* 18, 108. doi:10.1186/s13223-022-00744-5

Wark, P. a. B., Johnston, S. L., Moric, I., Simpson, J. L., Hensley, M. J., and Gibson, P. G. (2002). Neutrophil degranulation and cell lysis is associated with clinical severity in virus-induced asthma. *Eur. Respir. J.* 19, 68–75. doi:10.1183/09031936.02.00226302

Watanabe, H., Miyake, K., Matsuoka, T., Kojima, R., Sakurai, D., Masuyama, K., et al. (2021). *LPCAT2* methylation, a novel biomarker for the severity of cedar pollen allergic rhinitis in Japan. *Am. J. Rhinol&Allergy* 35, 631–639. doi:10.1177/1945892420983646

Wei, S., Lu, J., Lou, J., Shi, C., Mo, S., Shao, Y., et al. (2020). Gastric cancer tumor microenvironment characterization reveals stromal-related gene signatures associated with macrophage infiltration. *Front. Genet.* 11, 663. doi:10.3389/fgene.2020.00663

Weiss, A., and Attisano, L. (2013). The TGFbeta superfamily signaling pathway. *Wiley Interdiscip. Rev. Dev. Biol.* 2, 47–63. doi:10.1002/wdev.86

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. (Camb)* 2, 100141. doi:10.1016/j.xinn.2021.100141

Wu, X.-L., Li, R., Zhang, H.-W., Jin, R., Wang, J.-Y., Juan, C.-X., et al. (2018). Methylation status of ORMDL3 regulates cytokine production and p-ERK/MMP9 pathway expression. *Exp. Cell Res.* 372, 43–51. doi:10.1016/j.yexcr.2018.09.008

Xie, Z.-J., Guan, K., and Yin, J. (2019). Advances in the clinical and mechanism research of pollen induced seasonal allergic asthma. *Am. J. Clin. Exp. Immunol.* 8, 1–8.

Yang, Z., Wen, P., Chen, J., Kang, J., Xiang, Y., Ding, S., et al. (2022). DNA methylation regulatory patterns and underlying pathways behind the co-pathogenesis of allergic rhinitis and chronic spontaneous urticaria. *Front. Immunol.* 13, 1053558. doi:10.3389/fimmu.2022.1053558

Yu, G., and He, Q.-Y. (2016). ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* 12, 477–479. doi:10.1039/C5MB00663E

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi:10.1089/omi.2011.0118

Zhang, M., Wang, Y., Wang, Y., Jiang, L., Li, X., Gao, H., et al. (2020). Integrative analysis of DNA methylation and gene expression to determine specific diagnostic biomarkers and prognostic biomarkers of breast cancer. *Front. Cell Dev. Biol.* 8, 529386. doi:10.3389/fcell.2020.529386

Zhang, Y., Lan, F., and Zhang, L. (2021). Advances and highlights in allergic rhinitis. *Allergy* 76, 3383–3389. doi:10.1111/all.15044

Zheng, Z., Lai, C., Li, W., Zhang, C., Ma, K., and Yao, Y. (2021). Identification of a novel glycolysis-related LncRNA signature for predicting overall survival in patients with bladder cancer. *Front. Genet.* 12, 720421. doi:10.3389/fgene.2021.720421

Zhou, X., Franklin, R. A., Adler, M., Carter, T. S., Condiff, E., Adams, T. S., et al. (2022). Microenvironmental sensing by fibroblasts controls macrophage population size. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2205360119. doi:10.1073/pnas.2205360119

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 18 September 2024
DOI 10.3389/fgene.2024.1342205

OPEN ACCESS

EDITED BY
Wenjie Liu,
Sun Yat-sen University, China

REVIEWED BY
Manoj Khokhar,
All India Institute of Medical Sciences Jodhpur, India
Ricardo Pujol Borrell,
Autonomous University of Barcelona, Spain

*CORRESPONDENCE
Chiaw-Ling Chng,
chng.chiaw.ling@singhealth.com.sg

RECEIVED 21 November 2023
ACCEPTED 28 August 2024
PUBLISHED 18 September 2024

CITATION
Chng C-L, Lai OF, Seah L-L, Yong K-L, Chung YH-W, Goh R and Lim CK (2024) A combined transcriptomics and proteomics approach reveals S100A4 as a potential biomarker for Graves' orbitopathy. *Front. Genet.* 15:1342205. doi: 10.3389/fgene.2024.1342205

A combined transcriptomics and proteomics approach reveals S100A4 as a potential biomarker for Graves' orbitopathy

Chiaw-Ling Chng[1]*, Oi Fah Lai[2], Lay-Leng Seah[3], Kai-Ling Yong[3], Yvonne Hsi-Wei Chung[3], Rochelle Goh[2] and Che Kang Lim[2]

[1]Department of Endocrinology, Singapore General Hospital, Singapore, Singapore, [2]Department of Clinical and Translational Research, Singapore General Hospital, Singapore, Singapore, [3]Oculoplastic Department, Singapore National Eye Centre, Singapore, Singapore

Background: There are no reliable biomarkers to identify Graves' disease patients who will develop severe Graves' orbitopathy (GO). We hypothesize that integrating various omics platforms can enhance our understanding of disease mechanisms and uncover potential biomarkers. This study aimed to (1) elucidate the differential gene expression profile of orbital fibroblasts in GO during early adipogenesis to better understand disease mechanisms and (2) compare tear protein profiles from our earlier study and the transcriptome profiles of orbital fibroblasts (OFs) to identify possible biomarkers of the disease.

Methods: OFs were grown from orbital adipose tissue obtained from nine GO patients (three for discovery and six for validation experiments). Total RNA was extracted from OFs on day 0 as the baseline for each sample and from differentiated OFs on days 4 and 8. Protein–protein interaction (PPI) analysis and functional enrichment analysis were also carried out. The differentially expressed genes (DEGs) from the RNA sequencing experiments were then compared to the full tear proteome profile from the author's previous study, which examined the tear protein changes of GO patients based on fold change > 1.6 or < –1.6. FDR < 0.05 was applied within all datasets. Further validation of S100 calcium-binding protein A4 (S100A4) downregulation in GO was performed via quantitative real-time PCR (qPCR).

Results: The whole transcriptomic analysis revealed 9 upregulated genes and 15 downregulated genes in common between the discovery and validation experiments. From the PPI network analysis, an interaction network containing six identified DEGs (ALDH2, MAP2K6, MT2A, SOCS3, S100A4, and THBD) was observed. The functional enrichment network analysis identified a set of genes related to oxysterol production. S100A4 was found to be consistently downregulated in both our transcriptome studies and the full-tear proteome profile from the author's previous study.

Conclusion: Our study identified several DEGs and potential gene pathways in GO patients, which concurred with the results of other studies. Tear S100A4 may

serve as a biomarker for the propensity to develop thyroid eye disease (TED) in patients with autoimmune thyroid disease (AITD) before clinical manifestation and should be confirmed in future studies.

KEYWORDS

Graves' ophthalmology, Graves' orbitopathy, transcriptome, tear biomarkers, S100A4

Introduction

Graves' orbitopathy (GO) is the most common extrathyroidal manifestation of Graves' disease (GD), affecting up to half the patients, with 5% progressing to sight-threatening disease (Bahn, 2010). Despite advances in treatment, a significant proportion of patients are undiagnosed, resulting in long-term debilitating consequences, such as diplopia and exposure keratopathy, severely impairing vision and quality of life. Hence, early diagnosis and treatment form the cornerstone of the management of GO. Currently, there are no reliable biomarkers to identify GD patients who will develop severe GO. Orbital fibroblasts (OFs) play a pivotal role in the pathogenesis of GO, functioning both as a major target of inflammatory cytokines released by infiltrating immune cells and as active participants in the perpetuation of orbital disease (Dik et al., 2016). Orbital tissue expansion and fibrosis, central consequences of the pathogenic processes, result from the proliferation, extracellular matrix production, and differentiation of OFs into adipocytes and myofibroblasts. The molecular mechanisms underlying orbital adipogenesis in GO are not well understood, and several studies have explored transcriptome profiling via RNA sequencing during orbital fibroblast adipogenesis in GO patients to elucidate potential treatment targets (Lee et al., 2018; Kim et al., 2021; Bai et al., 2022). Recent studies have also shown the involvement of lacrimal glands in the pathogenesis of ocular surface damage in GO (Huang et al., 2014; Eckstein et al., 2004). Tear sampling provides a convenient and non-invasive method of analyzing an accessible body fluid for the investigation of potential biomarkers in ocular diseases. Previously, the author demonstrated different tear profiles in patients with different severity levels of GO, which suggests the role of the tear proteome as a potential biomarker of GO development and progression in patients with autoimmune thyroid disease (AITD) (Chng et al., 2018). This study aimed to (1) investigate the transcriptome profiles of orbital fibroblasts during early adipogenesis and (2) compare tear protein profiles from our earlier study and the transcriptome profiles of orbital fibroblasts to identify possible common molecular markers of the disease.

Methodology

Patient recruitment

The study was approved by the SingHealth Centralized Institutional Review Board. A total of nine patients with GO who underwent orbital decompression or eyelid mullerectomy (one case) were recruited for this study. All patients with GO were managed by an ophthalmologist. The diagnosis of GO was made based on diagnostic criteria defined by Bartley and Gorman (1995), i.e., GO is present if eyelid retraction occurs in association with thyroid dysfunction, exophthalmos, optic nerve dysfunction, or extraocular muscle involvement, and other confounding causes such as idiopathic orbital inflammation are excluded. All the patients with GO were euthyroid and had inactive disease at the time of surgery. Subjects who were smokers, diabetic, and had a recent (<3 months) intake of steroids, immunomodulatory agents, or orbital radiation were also excluded. Smoking is known to increase the incidence and severity of thyroid eye disease (TED) and induce numerous gene expression changes (Bahn, 2010). Smoking also affects the adipogenesis of orbital fibroblasts (Chng et al., 2014). Steroids, diabetes, immunomodulatory agents, and orbital radiation are potential confounding factors that may influence adipogenesis, inflammation, and overall transcriptional activity in orbital adipose tissues. For the discovery phase, orbital adipose tissue from three GO cases (age 48–68 years, all female patients) recruited between 2016 and 2017 was used for the experiments (Table 1). For the validation phase, orbital adipose tissue from six GO cases (age 27–61 years, five female and one male patient) recruited between 2020 and 2022 was used (Table 1).

Orbital fibroblast cultures

OFs were grown from the orbital adipose tissue obtained. The excised fat was minced into 0.1 cm × 0.1 cm pieces and placed in plastic culture flasks in M199 (Gibco, Life Technologies) supplemented with 10% fetal bovine serum (FBS), gentamycin (20 ug/mL), and penicillin/streptomycin (100 u/mL), allowing OFs to emerge from the tissue as described previously (Valyasevi et al., 1999). Fibroblasts were expanded over a period of 3–4 weeks till sufficient cells were obtained. When sufficient cells were available for inducing lipid accumulation, orbital fibroblasts were seeded at a density of 2.1×10^6 cells per mL into T75 culture flasks. When the cells approached confluence, an adipogenic induction medium (AIM) comprising IMDM supplemented with 10% FBS, 0.528 mM IBMX, 0.033 uM biotin, 0.001 mM dexamethasone, 0.2 mM indomethacin, and 0.174 uM insulin was used to induce lipid accumulation in orbital fibroblasts. The medium was changed every 2 days to an adipogenic maintenance medium (IMDM with 10% FBS supplemented with 0.174 uM insulin). The cycling of induction and maintenance media was carried out until D18. Cells differentiated until D18 in 6-well culture flasks were stained with Oil Red O on D18 to visualize lipid accumulation and adipocyte morphology (Figure 1). For this study, cultures from D0 to D4 and D0 to D8 were used for both the discovery and

10.3389/fgene.2024.1342205

TABLE 1 Clinical characteristics of cases recruited for the study.

Clinical information	Discovery set			Validation set					
	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Age (years)	48	68	66	27	61	40	33	30	60
Gender	F	F	F	F	F	F	M	F	F
GD duration (years)	10	0.5	9.5	4	25	4	0[a]	2	3
TED duration (years)	10	0.5	9.5	3	1	3	7	1.5	2
Main treatment for thyrotoxicosis	RAI	CMZ	CMZ	CMZ	T	RAI	Nil	CMZ	RAI
Prior treatment for TED	Nil	S	Nil	S	S	S	Nil	Nil	S
Medical hx	HTN	Nil	HTN/HLD	Nil	Nil	Nil	Nil	Nil	HTN/HLD
Dysthyroid optic neuropathy	No	No	No	No	No	No	No	No	No
CAS (0–10)	0	0	2	0	2	1	0	0	0
Surgery	BOD	ROD	BOD	BOD	ROD	EM	BOD	LOD	LOD
TRAB (0–1.5 IU/L)	0.5	4.3	12.8	>40	4.5	26.4	2.02	37.1	16

[a]Case 4 (validation set) is a euthyroid GO case.
RAI, radioiodine; CMZ, carbimazole; T, thyroidectomy; S, steroid; HTN, hypertension; HLD, hyperlipidemia; BOD, bilateral orbital decompression; ROD, right-orbital decompression; EM, eyelid mullerectomy; LOD, left-orbital decompression; TRAb, TSH receptor antibody.

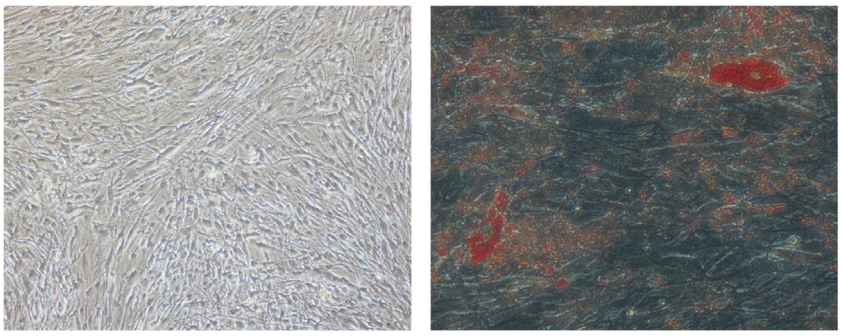


FIGURE 1
Digital photographs (at × 20 magnification) of OFs from GO subjects showing adipogenesis. (Left) D18 undifferentiated OFs. (Right) D18 differentiated OFs. The red spots are intracellular lipid droplets stained with Oil Red O.

validation phases. These time points were chosen to focus on the gene expression related to early adipogenesis.

Whole-transcriptome sequencing (total RNA sequence)

Total RNA was extracted from OFs on day 0 as the baseline for each sample and from differentiated OFs on days 4 and 8 using an RNeasy Kit (QIAGEN). Total RNA libraries were prepared using the Illumina TruSeq Kit and sequenced on the Illumina NovaSeq 6000 Sequencing Platform (Illumina, San Diego, CA, United States). Paired-end reads (150 bp) were mapped to the GRCh38 reference human genome. Output BCL files were converted to FASTQ files and demultiplexed using bcl2fastq v.2.20 from Illumina. Overall, approximately 40 million reads per sample were obtained.

Comparative tear protein samples

The differentially expressed genes (DEGs) from the RNA sequencing experiments were then compared to the full tear proteome profile from the author's previous studies, which examined the tear protein changes of GO patients with increasing severity compared to normal controls (Chng et al., 2018).

Quantitative real-time PCR for S100A4

Total RNA was extracted from OFs on day 0 as the baseline for each sample and from differentiated OFs from patients with GO on days 4 and 8 using the RNeasy Mini RNA Isolation and QIAshredder Kits (QIAGEN). A 15-min on-column DNase digestion was carried out to eliminate genomic DNA as

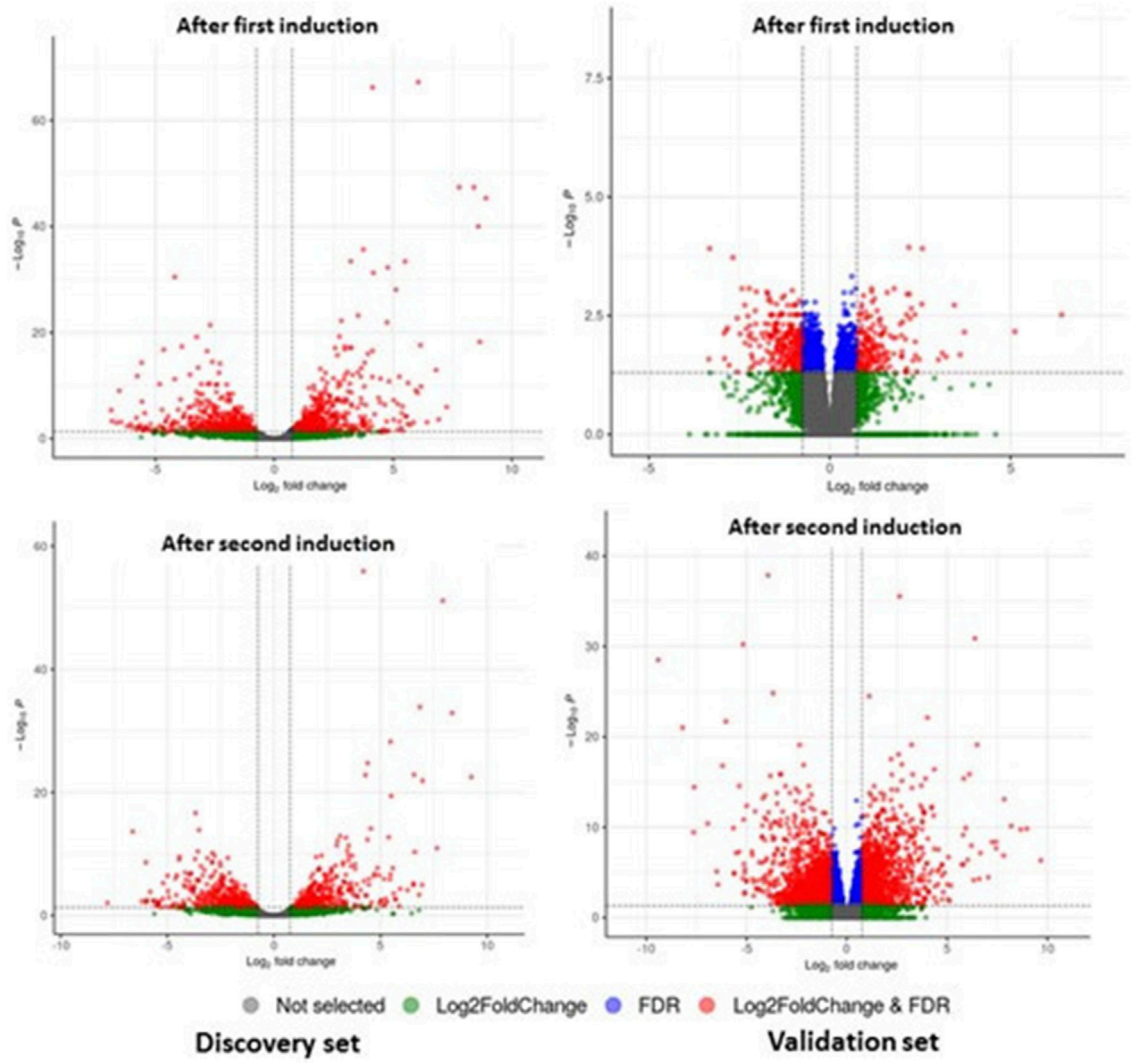


FIGURE 2
DEGs of mRNA between patients with GO and controls from discovery and validation phase experiments for first induction (D0–D4) and second induction (D0–D8). The volcano plots display the fold changes and *p*-values of differential mRNA expression in patients with GO compared to normal controls. Based on the relationship between fold change and statistical significance, subsets of mRNAs were isolated. The cut-off for logFC is 1.6. The red points represent the upregulated mRNAs with statistical significance $p < 0.05$, while the green dots represent the significantly downregulated expression.

outlined in the kit manual. Then, 1 ug of RNA was reverse-transcribed using a High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, United States). Quantitative real-time PCR (qPCR) was performed in duplicate using reagents from the GoTaq qPCR Kit (Promega) in 96-well hard-shell PCR plates (Bio-Rad). The primers used for S100 calcium-binding protein A4 (S100A4) analysis were as follows: forward primer, TCTTGGTTTGATCCTGACTGCT; reverse primer, ACTTGTCACCCTCTTTGCCC. Cycling conditions were induced using the Bio-Rad Real-Time Thermal Cycler. qPCR cycling was carried out according to the manufacturer's guidelines: 2 min of initial activation at 95°C, 40 cycles of denaturation at 95°C for 15 s, and annealing and extension for 1 min at 60°C. The relative gene expression of S100A4 was analyzed using the comparative CT ($2^{-\hat{o}\hat{o}\ CT}$) normalized against housekeeping gene β-actin.

Bioinformatics and statistical analysis

Sequencing read quality was assessed using FASTQC software, and low-quality bases were removed from individual reads using the cutadapt tool. Reads were mapped to genes through alignment to the human reference transcriptome (GENCODE version 38), and gene expression levels were quantified using Salmon Software. Differential gene expression analysis was conducted using DESeq2 (Love et al., 2014) by comparing baseline (D0) versus "after first induction" (D0–D4) and baseline (D0) versus "after second induction" (D0–D8). The statistical significance of differences was assessed using the Student's *t*-test. Differentially expressed mRNAs were identified by fold change (FC) > 1.6 or FC < −1.6 and false discovery rate (FDR) < 0.05 (Figure 2).

In addition to the gene-level analysis, protein–protein interaction (PPI) analysis and functional enrichment analysis

TABLE 2 DEGs in orbital fibroblasts derived from patients with GO when compared to normal controls. OFs were induced from D0 to D4 and from D0 to D8 for both discovery and validation phase experiments. FDR < 0.05; log2FC > 0.678 or < −0.678 (FC > 1.6 or < −1.6).

Gene	Gene description	Discovery set				Validation set			
		D0 to D4		D0 to D8		D0 to D4		D0 to D8	
		Log_2FC	FDR	Log_2FC	FDR	Log_2FC	FDR	Log_2FC	FDR
PDK4	Pyruvate dehydrogenase kinase 4	7.785	3.965E-48	7.920	7.459E-52	1.748	9.271E-03	6.110	1.363E-16
MAP2K6	Mitogen-activated protein kinase 6	5.036	5.862E-08	4.221	2.448E-06	1.396	5.048E-03	4.361	4.049E-17
HSD11B1	Hydroxysteroid 11-beta dehydrogenase 1	3.650	2.597E-08	3.811	2.556E-08	2.130	4.135E-02	4.865	1.068E-05
SLC19A2	Solute carrier family 19 member 2	3.043	9.190E-08	3.366	2.493E-07	1.396	3.663E-03	3.292	1.480E-16
MT2A	Metallothionein 2A	3.036	3.369E-08	3.237	2.302E-09	1.986	8.290E-03	3.467	2.111E-08
SLC16A9	Solute carrier family 16 member 9	2.612	2.539E-04	1.965	1.154E-02	1.202	2.806E-02	1.786	6.440E-06
GPAM	Glycerol-3-phosphate acyltransferase, mitochondrial	2.336	1.135E-08	2.125	4.487E-05	0.684	6.285E-03	2.785	3.231E-11
SAT1	Spermidine/spermine N1-acetyltransferase 1	2.034	2.340E-09	2.129	3.027E-08	1.135	3.242E-03	3.244	4.487E-15
ALDH2	Aldehyde dehydrogenase 2 family (mitochondrial)	1.415	3.725E-04	1.611	1.412E-04	0.767	7.473E-03	2.375	3.982E-12
SFRP4	Secreted frizzled related protein 4	−1.230	7.598E-04	−2.327	4.437E-04	−2.212	1.340E-02	−1.950	3.100E-02
AMACR	Alpha-methylacyl-CoA racemase	−1.668	8.859E-04	−1.465	1.804E-02	−0.729	4.990E-02	−0.678	4.501E-02
LRRN4CL	LRRN4 C-terminal-like	−1.895	2.703E-04	−2.006	1.456E-02	−0.863	6.527E-03	−1.318	3.826E-04
GALNT12	Polypeptide N-acetylgalactosaminyltransferase 12	−2.030	3.600E-03	−1.836	2.888E-02	−0.919	1.867E-02	−0.867	1.485E-03
IER2	Immediate early response 2	−2.154	4.163E-04	−1.942	2.854E-03	−0.834	4.135E-02	−1.361	5.404E-05
ARHGAP28	Rho GTPase activating protein 28	−2.158	6.247E-03	−2.367	9.543E-03	−1.257	1.378E-02	−1.397	1.675E-03
SOCS3	Suppressor of cytokine signaling 3	−2.188	1.231E-02	−2.776	1.114E-03	−0.886	1.308E-02	−1.728	7.384E-07
ECM2	Extracellular matrix protein 2	−2.400	1.472E-03	−1.768	4.824E-02	−1.731	1.032E-03	−2.335	2.957E-04
PSG1	Pregnancy-specific beta-1-glycoprotein 1	−2.711	2.877E-07	−3.619	2.070E-06	−2.214	4.089E-02	−5.081	4.356E-06
KCND3	Potassium voltage-gated channel subfamily D member 3	−2.721	5.700E-04	−2.492	1.345E-02	−1.598	4.283E-02	−2.116	2.117E-03
HSD3B7	Hydroxy-delta-5-steroid dehydrogenase, 3 beta-, and steroid delta-isomerase 7	−2.762	8.295E-04	−2.888	7.955E-04	−0.763	2.153E-02	−1.913	2.587E-06
EGR2	Early growth response 2	−2.809	3.338E-02	−2.814	4.636E-02	−3.344	2.599E-02	−3.930	3.478E-03
S100A4	S100 calcium-binding protein A4	−3.272	1.488E-09	−2.421	9.775E-06	−0.786	1.649E-02	−2.328	3.035E-07
THBD	Thrombomodulin	−3.616	1.379E-06	−4.135	1.593E-06	−1.103	3.009E-02	−2.492	6.021E-10
IER3	Immediate early response 3	−4.744	9.205E-03	−5.642	4.487E-03	−1.118	7.510E-03	−2.234	4.073E-08

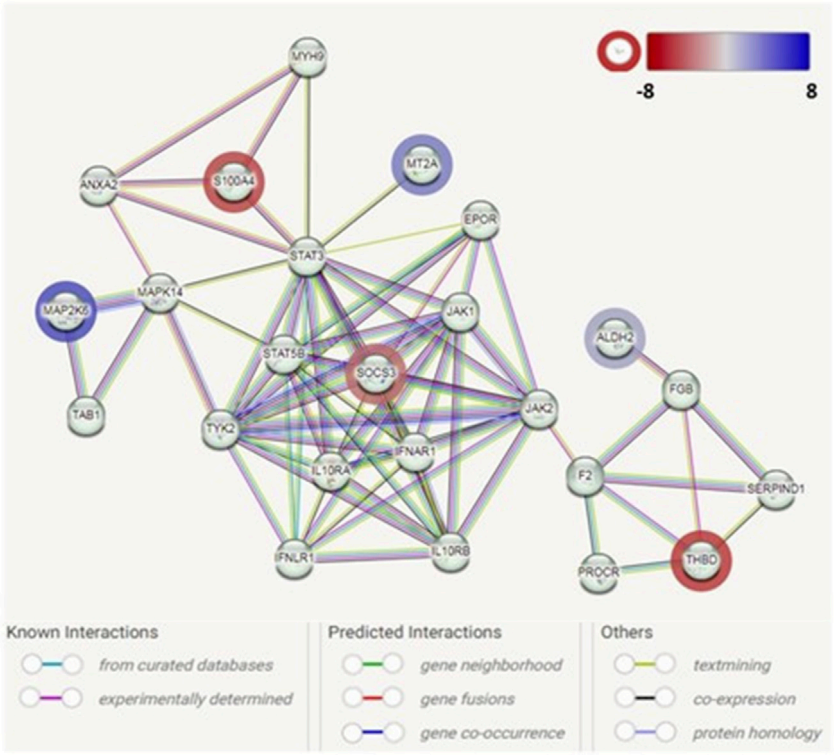


FIGURE 3
Protein–protein interaction network analysis among the identified genes associated with adipogenesis in GO. Each node represents a protein product. The node highlighted in red indicates downregulated genes/proteins. The node highlighted in blue indicates upregulated genes/proteins. Lines denote protein–protein interactions: light-blue lines represent interactions from curated databases; pink lines represent experimentally determined interactions; green lines represent interactions predicted by gene neighborhood; red lines represent interactions predicted by gene fusions; blue lines represent interactions predicted by gene co-occurrence; yellow lines represent interactions by text mining; black lines represent interactions by co-expression; and purple lines represent interactions by protein homology.

were carried out via Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database v12.0 (Szklarczyk et al., 2023) for the final enriched DEGs. Functional enrichment analysis was repeated using Database for Annotation, Visualization and Integrated Discovery (DAVID) (Sherman et al., 2022) to reconfirm the result.

Results

Whole-transcriptomic analysis identified 672 DEGs (FC ≥ 1.6 or FC < −1.6; FDR <0.05) in the discovery set from both induction cycles (D0–D4 and D0–D8) (Supplementary Table S1). On the other hand, 112 significantly differentially expressed genes were observed in the validation set (Supplementary Table S2) that were related to orbital adipogenesis. We next examined the shared (intersection) DEGs between 2 sets and found 24 DEGs during the early adipogenesis of OFs from GO patients after filtering with maintenance controls. The differential expression remained consistent in both the discovery and validation phase experiments. Of these, 9 genes were upregulated, and 15 were downregulated (Table 2).

PPI network associated with the early adipogenesis in GO

Based on the 24 DEGs listed as input, the main PPI network (Figure 3) associated with the early adipogenesis in GO containing the seed proteins and their neighbors' interactome was constructed. All the interactions between them were derived from all available active interaction sources in STRING. A medium-to-high level of confidence (score ≥ 0.50) was set. In addition, support from at least two different sources is required.

The PPI network analysis observed an interaction network containing six identified DEGs (*ALDH2*, *MAP2K6*, *MT2A*, *SOCS3*, *S100A4*, and *THBD*). The other nodes of the network mainly contain genes/proteins related to autoinflammatory/autoimmunity, such as JAK1, STAT3, STAT5B, IL10RA, and IL10RB, with *SOCS3* as a hub. Additionally, genes/protein nodes associated with lipid metabolism (e.g., F2 and FGB) were detected and connected with the detected DEG, *ALDH2*.

In addition, through the functional enrichment network analysis in the STRING database, we identified a set of genes—HSD11B1, HSD3B7, and AMACR—that were associated with pathways related to oxysterol production (FDR = 0.0248). The finding was further confirmed using DAVID (Fold enrichment was 31.20, *p*-value: 3.54E-03).

TABLE 3 Comparison of DEGs from the RNA sequencing experiments to the full tear proteome profile. FDR < 0.05; Log_2FC > 0.678 or < −0.678 (FC > 1.6 or < −1.6).

		Discovery set				Validation set				Tear proteome	
		After 1st cycle stimulation		After 2nd cycle stimulation		After 1st cycle stimulation		After 2nd cycle stimulation			
Gene	Gene description	Log_2FC	FDR	Log_2FC	FDR	Log_2FC	FDR	Log_2FC	FDR	Log_2FC	FDR
S100A4	S100 calcium-binding protein A4	−3.272	1.488E-09	−2.421	9.775E-06	−2.328	3.035E-07	−0.786	1.649E-02	−1.591	2.90E-02

Comparative transcriptomics and tear proteomic profile analysis

Next, a comparative transcriptomics and tear proteomic expression profile analysis was performed to identify potential tear marker(s) linked with GO. We compared the transcriptomics expression profile of all 68 tear proteins (24 up and 44 down) that were identified from TED/ GO patients previously (Chng et al., 2018). A standardized filtering cut-off, including consistent expression pattern, fold change >1.6 or < −1.6, and FDR < 0.05, was applied across all datasets (transcriptomics discovery sets, transcriptomics validation sets, and tear proteome sets). Based on the stringent filtering strategy, only one marker, S100A4, was found to be consistently downregulated (Table 3). The downregulation of S100A4 expression during early adipogenesis in orbital fibroblasts in GO was also confirmed by qPCR analysis in the current study, with a significantly reduced fold change in this gene expression on D4 and D8 compared to D0 fibroblast cultures while exposed to the adipogenic medium (Figure 4).

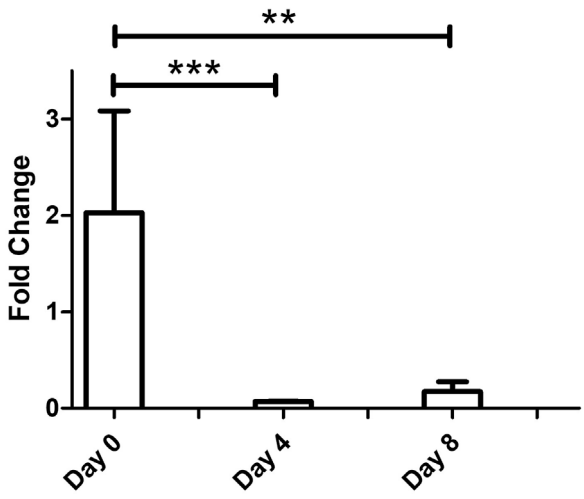


FIGURE 4
Downregulation of S100A4 relative gene expression in GO patients on days 0, 4, and 8, as measured by qPCR (*p < 0.05, **p < 0.01, and ***p < 0.001).

Discussion

Our study investigated DEG profiles during the early adipogenesis of OFs derived from patients with GO. These expression profiles were consistent between the discovery and validation experiments conducted using different cohorts of patients recruited over different time periods. Several of the gene expression patterns were noted in other studies related to GO. Pyruvate dehydrogenase kinase (PDK) enzymes trigger the switch from oxidative phosphorylation to cytoplasmic glycolysis, which plays a role in immune processes by promoting cell proliferation and strengthening antioxidant defense (Stacpoole, 2017). PDK2 overexpression has been found to promote perimysial orbital fibroblast proliferation in GO via Akt signaling (Ma et al., 2020). On the other hand, PDK4 has been shown to maintain the stability of HIF-1α (Ma et al., 2023). Notably, hypoxia-dependent HIF-1 activation was found to impact tissue remodeling in GO and may play a role in the worsening of GO due to smoking (Chng et al., 2014; Görtz et al., 2016).

Mitogen-activated protein (MAP2K6/MKK6), 11β-hydroxysteroid dehydrogenase 1 (11β-HSD1), and aldehyde dehydrogenase 2 (ALDH2) play important roles in adipogenesis. The expression of MKK6(Glu) in fibroblasts stimulates extensive p38-dependent adipocyte conversion in the absence of hormonal stimulation and strongly promotes adipogenesis both in 3T3-L1 fibroblasts and NIH-3T3 cells (Engelman et al., 1999). 11β-HSD1 is a bidirectional enzyme that interconverts inactive cortisone and active cortisol, thereby increasing local glucocorticoid bioavailability (Tomlinson et al., 2004). The induction of 11β-HSD1 activity and expression by inflammatory cytokines (TNF and IL-6) may enhance orbital adipogenesis and is pivotal in the regulation of local inflammatory response (Tomlinson et al., 2010). ALDH2 activation enhances adipogenesis and signaling pathways involving PPARγ and functions as a positive regulator of adipocyte differentiation (Yu et al., 2016). Cheng et al. (2018) showed that the expression of the anti-ALDH2 antibody was enhanced in GO patients and decreased in normal controls. In addition, anti-ALDH2 antibody levels are strongly correlated with the disease activity of GO (Cheng et al., 2018; Lin, 2024). The potential involvement of Wnt signaling dysregulation has been implicated in GO pathogenesis through the regulation of adipogenesis (Tao et al., 2017). The downregulation of secreted frizzled-related protein 4 (SFRP4), a secreted antagonist of the Wnt signaling pathway, was found in orbital fat from GO patients compared to normal controls (Ezra et al., 2012).

The protein encoded by spermidine/spermine N1 (SAT1) belongs to the acetyltransferase family and is a rate-limiting enzyme in the

catabolic pathway of polyamine metabolism. It catalyzes the acetylation of spermidine and spermine and is involved in the regulation of the intracellular concentration of polyamines and their transport out of cells. An increased concentration of spermine was found in the tears of patients with active GO compared to inactive GO (Billiet et al., 2022). The increased synthesis of spermine could be related to the overstimulation of Graves' autoantibodies and could play a proliferative role at the origin of the increased myogenesis and adipogenesis observed in GO. The suppressor of cytokine signaling (SOCS) proteins are key regulators of immune responses, and SOCS3 functions predominantly as a negative regulator of cytokines that activate the JAK-STAT3 pathway, which participates in the regulation of key biological processes, including cell proliferation, differentiation, and apoptosis (Jiang et al., 2018). SOCS-3 also binds to the insulin growth receptor 1 (IGF1R) and may be a direct substrate for the receptor tyrosine kinase. IGF1R pathways play a critical role in the pathogenesis of GO, and teprotumumab, a human monoclonal anti-IGF1R blocking antibody, has been approved by the FDA for the treatment of patients with GO, specifically in reducing proptosis (Kahaly et al., 2021).

Oxysterols are metabolites derived from cholesterol oxidation and have been implicated in the pathogenesis of several diseases (Poli et al., 2013). These metabolites bind to liver X receptors (LXRs) and retinoic acid receptor-related orphan receptors (RORs), which are members of the nuclear receptor family of transcription factors and modulate gene expression involved in inflammatory and autoimmune processes (Duc et al., 2019). RORγT is a transcription factor involved in Th17 cell development (Ivanov et al., 2006). Several studies have suggested the involvement of pathogenic Th17 cells in GO (Jiang et al., 2022). It is proposed that the interplay between Th17 cells and OFs promotes orbital inflammation and fibrosis, and it strengthens Th17-OF communication via augmented costimulatory molecules (Fang et al., 2017). Hypercholesterolemia is a novel risk factor for GO, with both total high- and low-density lipoprotein cholesterol levels having recently been found to be associated with the presence of GO (Sabini et al., 2018; Lanzolla et al., 2018). This association is supported by a phase-2 randomized clinical trial involving 88 patients with active moderate-to-severe GO and increased low-density lipoprotein cholesterol levels, which demonstrated that adding atorvastatin to intravenous glucocorticoids led to a better treatment outcome than when using intravenous glucocorticoids alone (Lanzolla et al., 2021). The link between hypercholesterolemia and GO may reflect an altered inflammatory state in hypercholesterolemia (Ridker and JUPITER Study Group, 2003; Fonseca and Izar, 2009). Hence, the effect of statin on GO may go beyond lowering cholesterol levels with its pleiotropic effects on the adipogenesis of OFs and immunomodulatory actions (Lanzolla et al., 2019). Taken together, we postulate that hypercholesterolemia contributes to the inflammatory milieu of GO via oxysterols. Statins, by lowering cholesterol levels, ameliorate the oxidative stress exerted by this downstream metabolite.

We found S100A4 downregulation in both our transcriptome studies and tear protein analyses in patients with GO. In our earlier study investigating differences in tear protein profiles in different stages of TED, we observed a downward trend of S100A4 fold change with increasing severity of TED in both discovery and verification phase experiments (Chng et al., 2018). It is interesting that the same trend is observed in our RNA sequencing data from orbital fibroblasts derived from patients with GO when compared to normal controls, which suggests a common pathophysiology. The downregulation of this gene expression in GO was also confirmed via qPCR experiments in the current study. In addition, an interaction between S100A4 and several other genes implicated in early adipogenesis was evident based on our PPI network analysis. Although we postulate that the dysregulation trend of this protein in the tears may reflect limbal normality, S100A4, which belongs to the S100 superfamily of small Ca^{2+}-binding proteins, also plays critical roles in the pathogenesis of autoimmune, fibrotic, and inflammatory disorders (Ambartsumian et al., 2019). Following the discovery of the correlation between the transcriptome and tear proteins, we explored protein–protein interaction networks between S100A4 and other candidate proteins based on our transcriptome analyses. Constructing protein interaction networks enables us to better understand disease mechanisms in GO through putative biological pathways. Such an approach has proven to be valuable in understanding the pathogenic mechanisms underlying other autoimmune diseases such as systemic lupus erythematosus (SLE), multiple sclerosis (MS), and type 1 diabetes (Safari-Alighiarloo et al., 2014). Notably, the candidate proteins from our PPI analysis revealed largely immune-mediated mechanisms with the involvement of the JAK-STAT3 pathway, interferon lambda receptor 1 and interferon alpha and beta receptor subunit 1 (IFNLR1 and IFNAR1), interleukin-10 receptor alpha and beta subunits (IL10RA and IL10RB, respectively), Annexin A2 (ANXA2), etc., some of which were found to be dysregulated in orbital tissues in patients with GO (Matheis et al., 2015; Falkowski et al., 2020; Gianoukakis et al., 2008; Wakelkamp et al., 2003). In addition, cross-interactions between S100A4 and proteins (based on our DEG results) previously noted to be dysregulated in GO, such as MAP2K6, SOCS3, and ALDH2, were also elucidated from this analysis.

There are limitations to our study. The small study sample size was due to strict inclusion and exclusion criteria for cases recruited for the study. Despite a small sample size, this study used next-generation sequencing technology and obtained robust results. In addition, the results of the discovery phase experiments were confirmed by the validation phase experiments, which increases the reliability of the results generated. Although the tear protein profile study was carried out at a different time point, the study was conducted in the same institution with similar patient profiles. Despite a small dataset, we observed a similar downward trend of S100A4 levels in both our transcriptome studies and tear protein analyses in patients with GO. This suggests that in the same disease milieu, whether at the orbital fibroblast or tear level, S100A4 potentially plays a role in the disease manifestations.

There were two patients with hyperlipidemia on statin treatment recruited for the study (one in the discovery phase and one in the validation phase), and recent studies suggest that statin may have beneficial effects on TED. Whether this has an implication on adipogenesis or gene expression is unknown. Although it may be difficult to control all lifestyle factors, we acknowledge that comorbidities, medication history, or lifestyle factors are potential confounders for our results and should be further explored in future studies.

Recent studies have demonstrated the utility of single-cell RNA sequencing (scRNA-seq) as a powerful tool to study cellular

heterogeneity and complex cellular events. Notably, Li et al. (2022) utilized scRNA-seq to create comprehensive transcriptional atlases of the cellular components in the orbital connective tissue (OCT) from healthy controls (HCs) and individuals with GO although several potential limitations still exist. Epigenetic profiling using single-cell ATAC-seq could also provide further resolution of the transcriptomic patterns found in our study. Although these techniques were not explored in our study, they would have been helpful in validating our findings.

Conclusion

The value of the transcriptome approach utilized in our study lies in its ability to generate candidate genes and pathways for further studies. Our study identified several DEGs and potential gene pathways in GO patients, which concurred with the results of other studies. This not only increased our understanding of the disease pathogenesis but also elucidated potential drug targets (e.g., teprotumumab and statins) and biomarkers (e.g., anti-ALDH2 antibody). The results of the PPI network analysis propose a biological explanation of this phenomenon, with potential crosstalk with other gene pathways discovered by our transcriptome analysis. Importantly, our study also identified S100A4 downregulation in both transcriptome studies and tear proteome analyses. Tear S100A4 may serve as a biomarker for the propensity to develop TED in patients with AITD before clinical manifestation and should be confirmed in future studies.

Data availability statement

The data presented in the study are deposited in the European Nucleotide Archive (ENA) repository, accession number PRJEB73574. The data can be found: https://www.ebi.ac.uk/ena/browser/view/PRJEB73574.

Ethics statement

The studies involving humans were approved by the SingHealth Centralized Institutional Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

C-LC: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, writing–original draft, and writing–review and editing. OFL: conceptualization, data curation, formal analysis, investigation, methodology, resources, writing–original draft, and writing–review and editing. L-LS: investigation, methodology, resources, and writing–review and editing. K-LY: investigation and writing–review and editing. YC: investigation and writing–review and editing. RG: investigation and writing–review and editing. CKL: data curation, formal analysis, investigation, methodology, software, validation, writing–original draft, and writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was supported by the SingHealth Duke-NUS AMC Academic Research Grant, official grant number: AM/SU050/2020.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2024.1342205/full#supplementary-material

References

Ambartsumian, N., Klingelhöfer, J., and Grigorian, M. (2019). "The multifaceted S100A4 protein in cancer and inflammation," in *Calcium-binding Proteins of the EF-hand superfamily. Methods in molecular biology*. Editor C. W. Heizmann (New York, NY: Springer New York), 339–365. doi:10.1007/978-1-4939-9030-6_22

Bahn, R. S. (2010). Graves' ophthalmopathy. *N. Engl. J. Med.* 362, 726–738. doi:10.1056/NEJMra0905750

Bai, X. J., Chu, X. J., Yang, T., Tong, B. D., Qi, X., Peng, Y. Y., et al. (2022). Changes of mRNA expression underlying orbital adipogenesis in thyroid-associated orbitopathy. *Arch. Med. Sci.* 18, 1708–1715. doi:10.5114/aoms/153478

Bartley, G. B., and Gorman, C. A. (1995). Diagnostic criteria for Graves' ophthalmopathy. *Am. J. Ophthalmol.* 119, 792–795. doi:10.1016/s0002-9394(14)72787-4

Billiet, B., Chao de la Barca, J. M., Ferré, M., Muller, J., Vautier, A., Assad, S., et al. (2022). A tear metabolomic profile showing increased ornithine decarboxylase activity and spermine synthesis in thyroid-associated orbitopathy. *J. Clin. Med.* 11, 404. doi:10.3390/jcm11020404

Cheng, K.-C., Wu, Y.-J., Cheng, K.-H., Cheng, K.-Y., Chen, K.-J., Wu, W.-C., et al. (2018). Autoantibody against aldehyde dehydrogenase 2 could be a biomarker to monitor progression of Graves' orbitopathy. *Graefes Arch. Clin. Exp. Ophthalmol.* 256, 1195–1201. doi:10.1007/s00417-017-3894-4

Chng, C. L., Lai, O. F., Chew, C. S.-M., Peh, Y. P., Fook-Chong, S. M.-C., Seah, L. L., et al. (2014). Hypoxia increases adipogenesis and affects adipocytokine production in orbital fibroblasts-a possible explanation of the link between smoking and Graves' ophthalmopathy. *Int. J. Ophthalmol.* 7, 403–407. doi:10.3980/j.issn.2222-3959.2014.03.03

Chng, C.-L., Seah, L. L., Yang, M., Shen, S. Y., Koh, S. K., Gao, Y., et al. (2018). Tear proteins calcium binding protein A4 (S100A4) and prolactin induced protein (PIP) are potential biomarkers for thyroid eye disease. *Sci. Rep.* 8, 16936. doi:10.1038/s41598-018-35096-x

Dik, W. A., Virakul, S., and van Steensel, L. (2016). Current perspectives on the role of orbital fibroblasts in the pathogenesis of Graves' ophthalmopathy. *Exp. Eye Res.* 142, 83–91. doi:10.1016/j.exer.2015.02.007

Duc, D., Vigne, S., and Pot, C. (2019). Oxysterols in autoimmunity. *Int. J. Mol. Sci.* 20, 4522. doi:10.3390/ijms20184522

Eckstein, A. K., Finkenrath, A., Heiligenhaus, A., Renzing-Köhler, K., Esser, J., Krüger, C., et al. (2004). Dry eye syndrome in thyroid-associated ophthalmopathy: lacrimal expression of TSH receptor suggests involvement of TSHR-specific autoantibodies. *Acta Ophthalmol. Scand.* 82, 291–297. doi:10.1111/j.1395-3907.2004.00268.x

Engelman, J. A., Berg, A. H., Lewis, R. Y., Lin, A., Lisanti, M. P., and Scherer, P. E. (1999). Constitutively active mitogen-activated protein kinase kinase 6 (MKK6) or salicylate induces spontaneous 3T3-L1 adipogenesis. *J. Biol. Chem.* 274, 35630–35638. doi:10.1074/jbc.274.50.35630

Ezra, D. G., Krell, J., Rose, G. E., Bailly, M., Stebbing, J., and Castellano, L. (2012). Transcriptome-level microarray expression profiling implicates IGF-1 and Wnt signalling dysregulation in the pathogenesis of thyroid-associated orbitopathy. *J. Clin. Pathol.* 65, 608–613. doi:10.1136/jclinpath-2012-200719

Falkowski, B., Szczepanek-Parulska, E., Sawicka-Gutaj, N., Krygier, A., and Ruchala, M. (2020). Evaluation of IL-29 in euthyroid patients with Graves' orbitopathy: a preliminary study. *Mediat. Inflamm.* 2020, 4748612. doi:10.1155/2020/4748612

Fang, S., Huang, Y., Zhong, S., Li, Y., Zhang, Y., Li, Y., et al. (2017). Regulation of orbital fibrosis and adipogenesis by pathogenic Th17 cells in Graves orbitopathy. *J. Clin. Endocrinol. Metabolism* 102, 4273–4283. doi:10.1210/jc.2017-01349

Fonseca, F. A. H., and Izar, M. C. O. (2009). Primary prevention of vascular events in patients with high levels of C-reactive protein: the JUPITER study. *Expert Rev. Cardiovasc Ther.* 7, 1041–1056. doi:10.1586/erc.09.93

Gianoukakis, A. G., Khadavi, N., and Smith, T. J. (2008). Cytokines, Graves' disease, and thyroid-associated ophthalmopathy. *Thyroid* 18, 953–958. doi:10.1089/thy.2007.0405

Görtz, G.-E., Horstmann, M., Aniol, B., Reyes, B. D., Fandrey, J., Eckstein, A., et al. (2016). Hypoxia-dependent HIF-1 activation impacts on tissue remodeling in Graves' ophthalmopathy-implications for smoking. *J. Clin. Endocrinol. Metab.* 101, 4834–4842. doi:10.1210/jc.2016-1279

Huang, D., Luo, Q., Yang, H., and Mao, Y. (2014). Changes of lacrimal gland and tear inflammatory cytokines in thyroid-associated ophthalmopathy. *Invest Ophthalmol. Vis. Sci.* 55, 4935–4943. doi:10.1167/iovs.13-13704

Ivanov, I. I., McKenzie, B. S., Zhou, L., Tadokoro, C. E., Lepelley, A., Lafaille, J. J., et al. (2006). The orphan nuclear receptor RORgammat directs the differentiation program of proinflammatory IL-17+ T helper cells. *Cell* 126, 1121–1133. doi:10.1016/j.cell.2006.07.035

Jiang, M., Zhao, J., Wang, P., Yan, S., and Wang, Y. (2022). Research progress in Th17 cells and the relevant cytokines in Graves' ophthalmopathy. *Zhong Nan Da Xue Xue Bao Yi Xue Ban.* 47, 1748–1753. doi:10.11817/j.issn.1672-7347.2022.220205

Jiang, W., Peng, Y., and Yang, K. (2018). Cellular signaling pathways regulating β-cell proliferation as a promising therapeutic target in the treatment of diabetes. *Exp. Ther. Med.* 16, 3275–3285. doi:10.3892/etm.2018.6603

Kahaly, G. J., Douglas, R. S., Holt, R. J., Sile, S., and Smith, T. J. (2021). Teprotumumab for patients with active thyroid eye disease: a pooled data analysis, subgroup analyses, and off-treatment follow-up results from two randomised, double-masked, placebo-controlled, multicentre trials. *Lancet Diabetes and Endocrinol.* 9, 360–372. doi:10.1016/S2213-8587(21)00056-5

Kim, D. W., Taneja, K., Hoang, T., Santiago, C. P., McCulley, T. J., Merbs, S. L., et al. (2021). Transcriptomic profiling of control and thyroid-associated orbitopathy (TAO) orbital fat and TAO orbital fibroblasts undergoing adipogenesis. *Invest Ophthalmol. Vis. Sci.* 62, 24. doi:10.1167/iovs.62.9.24

Lanzolla, G., Sabini, E., Leo, M., Menconi, F., Rocchi, R., Sframeli, A., et al. (2021). Statins for Graves' orbitopathy (STAGO): a phase 2, open-label, adaptive, single centre, randomised clinical trial. *Lancet Diabetes and Endocrinol.* 9, 733–742. doi:10.1016/S2213-8587(21)00238-2

Lanzolla, G., Sabini, E., Profilo, M. A., Mazzi, B., Sframeli, A., Rocchi, R., et al. (2018). Relationship between serum cholesterol and Graves' orbitopathy (GO): a confirmatory study. *J. Endocrinol. Invest* 41, 1417–1423. doi:10.1007/s40618-018-0915-z

Lanzolla, G., Vannucchi, G., Ionni, I., Campi, I., Sileo, F., Lazzaroni, E., et al. (2019). Cholesterol serum levels and use of statins in Graves' orbitopathy: a new starting point for the therapy. *Front. Endocrinol. (Lausanne)* 10, 933. doi:10.3389/fendo.2019.00933

Lee, B. W., Kumar, V. B., Biswas, P., Ko, A. C., Alameddine, R. M., Granet, D. B., et al. (2018). Transcriptome analysis of orbital adipose tissue in active thyroid eye disease using next generation RNA sequencing technology. *Open Ophthalmol. J.* 12, 41–52. doi:10.2174/1874364101812010041

Li, Z., Wang, M., Tan, J., Zhu, L., Zeng, P., Chen, X., et al. (2022). Single-cell RNA sequencing depicts the local cell landscape in thyroid-associated ophthalmopathy. *Cell Rep. Med.* 3, 100699. doi:10.1016/j.xcrm.2022.100699

Lin, C. C., Chiu, L. W., Lee, P. Y., Chen, K. J., and Cheng, K. C. (2024). Comparison of the correlation between serum autoantibodies against aldehyde dehydrogenase 2 and thyrotropin receptor in patients with Graves' ophthalmopathy. *Indian journal of ophthalmology* 72 (2), 298–299. doi:10.4103/IJO.IJO_673_23

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, 550. doi:10.1186/s13059-014-0550-8

Ma, R., Gan, L., Ren, H., Harrison, A., and Qian, J. (2020). PDK2-enhanced glycolysis promotes fibroblast proliferation in thyroid-associated ophthalmopathy. *J. Mol. Endocrinol.* 65, 163–174. doi:10.1530/JME-20-0143

Ma, Z., Mo, R., Yang, P., Ding, Y., Zhang, H., Dong, Z., et al. (2023). PDK4 facilitates fibroblast functions and diabetic wound healing through regulation of HIF-1α protein stability and gene expression. *FASEB J.* 37, e23215. doi:10.1096/fj.202300874RR

Matheis, N., Lantz, M., Grus, F. H., Ponto, K. A., Wolters, D., Brorson, H., et al. (2015). Proteomics of orbital tissue in thyroid-associated orbitopathy. *J. Clin. Endocrinol. Metab.* 100, E1523–E1530. doi:10.1210/jc.2015-2976

Poli, G., Biasi, F., and Leonarduzzi, G. (2013). Oxysterols in the pathogenesis of major chronic diseases. *Redox Biol.* 1, 125–130. doi:10.1016/j.redox.2012.12.001

Ridker, P. M., and JUPITER Study Group (2003). Rosuvastatin in the primary prevention of cardiovascular disease among patients with low levels of low-density lipoprotein cholesterol and elevated high-sensitivity C-reactive protein: rationale and design of the JUPITER trial. *Circulation* 108, 2292–2297. doi:10.1161/01.CIR.0000100688.17280.E6

Sabini, E., Mazzi, B., Profilo, M. A., Mautone, T., Casini, G., Rocchi, R., et al. (2018). High serum cholesterol is a novel risk factor for Graves' orbitopathy: results of a cross-sectional study. *Thyroid* 28, 386–394. doi:10.1089/thy.2017.0430

Safari-Alighiarloo, N., Taghizadeh, M., Rezaei-Tavirani, M., Goliaei, B., and Peyvandi, A. A. (2014). Protein-protein interaction networks (PPI) and complex diseases. *Gastroenterol. Hepatol. Bed Bench* 7, 17–31.

Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., and Lane, H. C. (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* 50, W216–W221.

Stacpoole, P. W. (2017). Therapeutic targeting of the pyruvate dehydrogenase complex/pyruvate dehydrogenase kinase (PDC/PDK) Axis in cancer. *J. Natl. Cancer Inst.* 109. doi:10.1093/jnci/djx071

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., et al. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* 51, D638–D646. doi:10.1093/nar/gkac1000

Tao, W., Ayala-Haedo, J. A., Field, M. G., Pelaez, D., and Wester, S. T. (2017). RNA-sequencing gene expression profiling of orbital adipose-derived stem cell population implicate HOX genes and WNT signaling dysregulation in the pathogenesis of thyroid-associated orbitopathy. *Invest Ophthalmol. Vis. Sci.* 58, 6146–6158. doi:10.1167/iovs.17-22237

Tomlinson, J. W., Durrani, O. M., Bujalska, I. J., Gathercole, L. L., Tomlins, P. J., Reuser, T. T. Q., et al. (2010). The role of 11beta-hydroxysteroid dehydrogenase 1 in adipogenesis in thyroid-associated ophthalmopathy. *J. Clin. Endocrinol. and Metabolism* 95, 398–406. doi:10.1210/jc.2009-0873

Tomlinson, J. W., Walker, E. A., Bujalska, I. J., Draper, N., Lavery, G. G., Cooper, M. S., et al. (2004). 11beta-hydroxysteroid dehydrogenase type 1: a tissue-specific regulator of glucocorticoid response. *Endocr. Rev.* 25, 831–866. doi:10.1210/er.2003-0031

Valyasevi, R. W., Erickson, D. Z., Harteneck, D. A., Dutton, C. M., Heufelder, A. E., Jyonouchi, S. C., et al. (1999). Differentiation of human orbital preadipocyte fibroblasts induces expression of functional thyrotropin receptor. *J. Clin. Endocrinol. Metab.* 84, 2557–2562. doi:10.1210/jcem.84.7.5838

Wakelkamp, I. M. M. J., Bakker, O., Baldeschi, L., Wiersinga, W. M., and Prummel, M. F. (2003). TSH-R expression and cytokine profile in orbital tissue of active vs. inactive Graves' ophthalmopathy patients. *Clin. Endocrinol. (Oxf)* 58, 280–287. doi:10.1046/j.1365-2265.2003.01708.x

Yu, Y.-H., Liao, P.-R., Guo, C.-J., Chen, C.-H., Mochly-Rosen, D., and Chuang, L.-M. (2016). PKC-ALDH2 pathway plays a novel role in adipocyte differentiation. *PLoS One* 11, e0161993. doi:10.1371/journal.pone.0161993

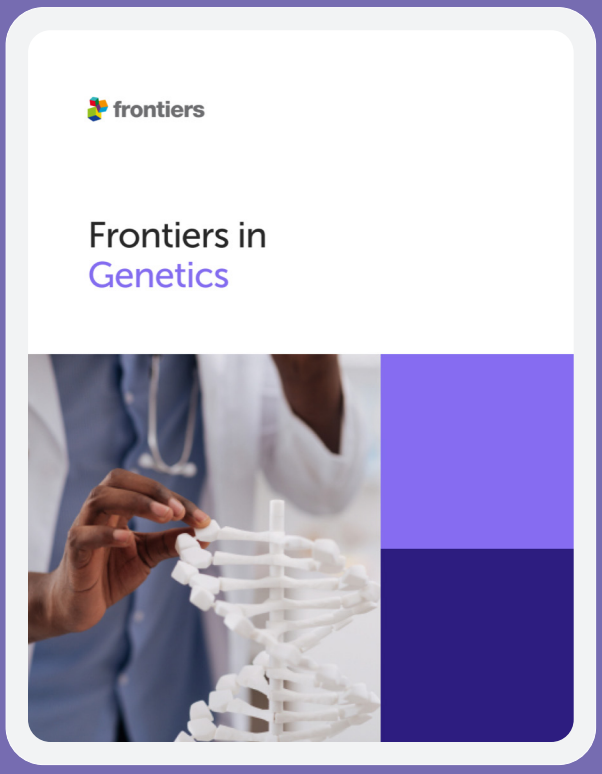
Frontiers in Genetics

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