Research article

Identification of therapeutic targets and medicines for comorbid Crohn's disease and rheumatoid arthritis: A comprehensive analysis

Qiu Dong [a,1], Xiaoting Wu [b,1], Tsz Mok [c], Gaohan Cai [a], Zhengang Zha [a], Guorong She [a,**], Junyuan Chen [a,*]

[a] *Center for Bone, Joint and Sports Medicine, The First Affiliated Hospital of Jinan University, Guang Zhou, 510665, China*
[b] *Guangdong Provincial Key Laboratory of Pharmaceutical Bioactive Substances, Guangdong Pharmaceutical University, Guang Zhou, 510665, China*
[c] *Department of Infectious Diseases and Public Health, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Hong Kong, 999077, China*

ARTICLE INFO

ABSTRACT

Objective: The incidence of Crohn's disease (CD) and rheumatoid arthritis (RA) co-morbidity, as well as the number of individuals affected, is on the rise due to their shared molecular and cellular factors. This study aimed to identify potential therapeutic targets and medicines for comorbid CD and RA.

Methods: We integrated single-cell RNA sequencing, Mendelian randomization, and colocalization analysis results from public databases to analyse immune cell subgroups in CD and RA patients and identify candidate therapeutic targets. We further screened potential medicines for the identified candidate targets using the Comparative Toxicogenomics Database (CTD) and molecular docking and molecular dynamics simulations.

Results: The proportion of CD8 effector memory T cells (Tem) was consistently elevated in the peripheral blood mononuclear cells (PBMCs) of both CD and RA patients. MYBL1 had a causal effect on the onset of both CD (OR = 1.23; 95 % CI, 1.05–1.45; P = 0.046) and RA (OR = 1.45; 95 % CI, 1.14–1.85; P = 0.04). Four potential therapeutic molecules were retrieved from the CTD database, among which tretinoin (docking score: −6.3 kcal/mol) showed the best potential.

Conclusion: Our comprehensive analysis suggests that CD8 Tem cells are a key cell group in comorbid RA and CD and that MYBL1 has a causal effect. Tretinoin was identified as a potential targeted therapeutic drug, which is of great clinical research value.

1. Introduction

Crohn's disease (CD) and rheumatoid arthritis (RA) are both chronic autoimmune diseases that can result in debilitating

Abbreviations (if any)

ANXA1   annexin A1
BH   Benjamini-Hochberg
CD   crohn's disease
CI   confidence interval
CT   control
CTD   comparative toxicogenomics database
CTSW   contrary cathepsin W
DC   dendritic cell
DEGs   differentially expressed genes
DMAIDs   anti-inflammatory bowel disease medicines
DMARDs   disease-modifying anti-rheumatic medicines
EBI   European Bioinformatics Institute
eQTL: expression quantitative trait loci
FKBP11   FK506 binding protein 11
GBP5   guanylate binding protein 5
GEO   gene expression omnibus
GO   gene ontology
GWAS   genome-wide association study
GZMB   granzyme B
GZMH   granzyme H
ITGA4   integrin subunit alpha 4
IVW   inverse variance-weighted
KEGG   Kyoto Encyclopedia of Genes and Genomes
KLRD1   lectin-like receptor D1
MIAT   myocardial infarction-associated transcript
MIF   migration inhibitory factor
MM/GBSA   molecular mechanics/generalized born surface area
MR   mendelian randomization
MRPL10   mitochondrial ribosomal protein L10
MYBL1   myb protooncogene like 1
NK   natural killer
NSAIDs   nonsteroidal anti-inflammatory medicines
OR   odds ratio
PBMCs   peripheral blood mononuclear cells
PCA   principal component analysis
PTPN4   protein tyrosine phosphatase nonreceptor type 4
RA   rheumatoid arthritis
RMSD   root mean square deviation
RMSF   root mean square fluctuation
RNA-seq   transcriptome sequencing
scRNA-seq single-cell RNA sequencing
SNP   single nucleotide polymorphism
SUN2   sad1 and UNC84 domain containing 2
Tef   effector T cells
Tem   effector memory T cells
TIGIT   T cell immunoreceptor with Ig and ITIM domains
Tn   naive T cells
TNFAIP3   tumor necrosis factor alpha-induced protein 3
UMAP   uniform manifold approximation and projection

complications affecting the gastrointestinal tract and joints, respectively [1,2]. Over the past 30 years, the incidence of both diseases has steadily increased with projections for continued rise [3,4]. These diseases share common pathogenic mechanisms, including dysregulated immune responses and genetic predisposition. Importantly, a growing body of research indicates a high likelihood of comorbidity between these two disorders. Studies have shown that patients with CD demonstrate a 3.35 times higher risk of developing RA compared to individuals without CD [5]. Meanwhile, RA is associated with an 11 % increased risk of CD [6]. The increasing prevalence and burden of this comorbidity highlight the urgent need for effective therapeutic strategies, which we aim to address in this study.

Presently, pharmaceutical interventions serve as the first line of defence against these two diseases, with clinical remission as the primary goal. In recent years, the use of anti-inflammatory bowel disease medicines (DMAIDs) and disease-modifying anti-rheumatic medicines (DMARDs) has shown favourable outcomes in the treatment of CD and RA [7,8]. However, in populations with comorbid CD and RA, there are limited safe and effective therapeutic alternatives. While certain immunosuppressants and biologic agents demonstrate therapeutic efficacy in both CD and RA, their long-term use may increase susceptibility to infections and tumors. Moreover, despite the significant potency of these biologic agents, they are expensive and also may trigger severe allergic reactions. Furthermore, some commonly used medicines for the treatment of RA, such as penicillamine, gold preparations, and nonsteroidal anti-inflammatory medicines (NSAIDs), may cause various gastrointestinal complications or even worsen CD [9].

Peripheral blood mononuclear cells (PBMCs) play a pivotal role in modulating the body's immune response. Interacting with PBMCs can effectively modulate the body's immune response and inhibit inflammatory responses, making them a highly promising therapeutic target. In recent years, an increasing number of studies have focused on investigating the significance of PBMCs in diagnosing, monitoring disease activity, and evaluating treatment response [4]. Pu demonstrated that the differentiation and development of distinct PBMC subpopulations are crucial for maintaining immunological homeostasis, suppressing inflammatory responses, and improving the prognosis of CD [10]. Recent studies have shown that monocytes in PBMCs can migrate to the synovial tissue in response to chemokines, differentiate into macrophages, and secrete proinflammatory factors that promote inflammatory responses in RA [11]. Therefore, conducting an in-depth examination of the role of PBMCs in CD and RA progression could potentially lead to the discovery of more effective interventions aimed at improving patients' quality of life.

While traditional methods including observational studies and animal models have been instrumental in disease research, they are often limited by confounding factors and lack the capability to address cell heterogeneity. Single-cell RNA sequencing (scRNA-seq) has emerged as a potent tool that enables an in-depth exploration of immune cell subpopulations within PBMCs, thereby unveiling the complexity of biological systems. When combined with scRNA-seq, several complementary technologies greatly help reveal the molecular mechanisms of complex diseases. Expression quantitative trait loci (eQTL) analysis aids in understanding the genetic regulation of gene expression, while Mendelian randomization helps to determine the causal relationships between genes and diseases [12,13].
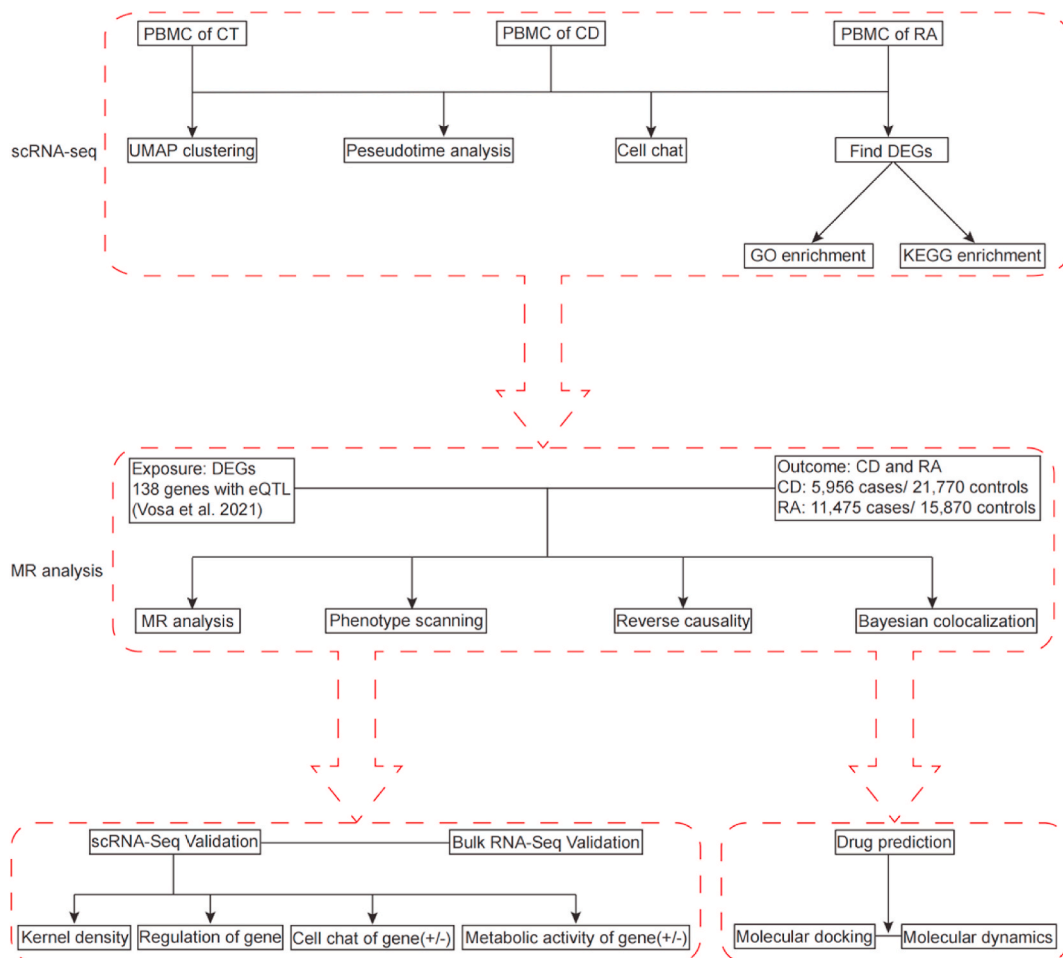

Fig. 1. Study design for identification of genes causally associated with CD and RA.

Colocalization analysis assists in identifying common genetic variants associated with multiple traits or diseases [14]. Transcriptome sequencing (RNA-seq) allows for a comprehensive analysis of gene expression patterns. Molecular docking and molecular dynamics simulation are helpful in studying the interactions between drug targets and ligands and the resulting conformational changes. Despite these advances, comprehensive understanding of comorbidities, particularly between CD and Rheumatoid RA, remains a significant gap in literature. This study addresses this issue by integrating advanced technologies to analyse PBMC sequencing data, aiming to identify potential therapeutic targets and treatments for comorbid CD and RA. The study workflow is illustrated in Fig. 1.

2. Materials and methods

2.1. Data source and preparation

The scRNA-seq data (10X Genomics data) from PBMC of CD (GSE163314) and RA (GSE159117) patients were initially retrieved and downloaded from the Gene Expression Omnibus (GEO) database. Detailed information about these datasets can be found in Table 1. To analyse the scRNA-seq data, an initial screening process was conducted using the Seurat R package [15]. During this screening, only genes expressed in a minimum of 3 cells and cells expressing at least 200 genes were retained. The PercentageFeatureSet function was utilized to assess the ratio of mitochondrial DNA to rRNA. Additionally, we selected cells with feature counts ranging from 100 to 4000 and excluded cells with mitochondrial gene counts exceeding 10 %. Ultimately, our analysis included a total of 35,282 single-cell data points, comprising 7428 from normal controls (CT), 17,690 from CD patients and 10,164 from RA patients.

The eQTL data used in this study are derived from an extensive dataset made publicly accessible by Vosa et al. [16]. The genetic variants that exhibited a substantial association with CD were obtained from a comprehensive genome-wide association study (GWAS) carried out by the European Bioinformatics Institute (EBI), involving 5956 CD patients and 14,927 controls (CTs) [17]. Similarly, the genetic variants significantly associated with RA were extracted from another large GWAS that was also conducted by the EBI and encompassed a cohort of 13,838 individuals diagnosed with RA and 33,742 CTs [18]. To enhance the credibility of our results, we performed an external validation utilizing GWAS provided by the University of Bristol, which comprised 17,897 CD patients and 33, 977 CTs [17].

2.2. Sequencing data analysis

First, we applied log normalization to normalize the gene expression of individual cells. Subsequently, we identified 2000 highly variable genes using the FindVariableFeatures function. We then performed principal component analysis (PCA) to downscale the data, effectively addressing batch effects between samples by applying the Harmony method. Next, we utilized the uniform manifold approximation and projection (UMAP) nonlinear dimensionality reduction technique and visualized the results through the DimPlot function. We repeated these steps after extracting T-cell clusters from PBMCs. We employed the Slingshot package to infer cell lineage and pseudotemporal relationships [19]. Additionally, the CellChat package was utilized to infer intercellular communication and identify significant pathways based on gene expression interactions, signalling ligands, receptors, and cofactors [20]. Then, highly differentially expressed genes (DEGs) (at least 2-fold) in CD8 Tem cells relative to other cells were identified using the FindMarkers function. Finally, GO and KEGG enrichment analysis of genes specifically highly expressed in the screened CD8 Tem cells was performed using the clusterProfiler package [21].

2.3. Mendelian randomization analysis and Bayesian colocalization analysis

In this study, we conducted MR analysis of the DEGs identified through single-cell analysis using the "TwoSampleMR" package. Only genes with significant eQTL associations ($P < 5 \times 10^{-8}$) were considered, using CD and RA as outcome indicators. Furthermore, phenotypic scanning was performed using the "phenoscanner", and SNPs associated with known risk factors for CD or RA were excluded. When a given gene contained only one eQTL, the Wald ratio was employed. Inverse variance-weighted MR (MR-IVW) was used as multiple genetic instrumental variables were present, and heterogeneity analysis was performed. SNPs in the exposure and outcome GWAS were directionally harmonized and Steiger filtered. Subsequently, common causative genes of CD and RA were selected for reverse causality analysis.

In addition, Bayesian colocalization analysis was employed using the "coloc" package to evaluate the correlation between the two traits. And, determined from previous articles, when PPH_4 > 80 % was defined as the evidence of colocalization of genes [12].

Table 1
Detailed information about the data.

ID	Type	Descriptive
GSM4976997	scRNA-seq	scRNA-seq for PBMC from healthy control
GSM4977001	scRNA-seq	scRNA-seq for PBMC from healthy control
GSM4976999	scRNA-seq	scRNA-seq for PBMC from CD patient
GSM4977003	scRNA-seq	scRNA-seq for PBMC from CD patient
GSE159117	scRNA-seq	scRNA-seq for PBMC from RA patient
ebi-a-GCST003044	GWAS	Including 5956 CD cases and 21,770 population controls of European descent with 110,583 SNPs.
ebi-a-GCST005569	GWAS	Including 11,475 RA cases and 15,870 controls of European descent with 129,464 SNPs.

2.4. Role of pathogenic genes at the scRNA-seq

To investigate the involvement of pathogenic genes in T cells, we initially conducted weighted kernel density estimation using the Nebulosa package. This analysis allowed us to determine the densities and weights of disease-causative genes across different T cell subpopulations. Subsequently, GeneSwitches was employed to calculate the switching probabilities of genes over pseudo time, which provided insights into the activation and repression patterns of genes over their developmental trajectories. Based on the above analyses, T cell subpopulations exhibiting elevated levels of pathogenic genes in both CD and RA were classified according to the switching behaviour of the pathogenic genes as positive gene populations or negative cell populations. Subsequently, the intercellular communication and metabolic activities between these cell populations were further investigated with the aim of revealing their functional interactions.

2.5. Drug prediction

We used MYBL1, CD, and RA as keywords to construct the disease–gene–drug linkage network using the Comparative Toxicogenomics Database (CTD) and screened for drugs targeting candidates that play key roles in disease development [22]. Furthermore, the structures of small molecule drugs and targets were retrieved from the PDB and AlphaFold protein structure database, and molecular docking between the two was performed using AutoDock software [23,24]. Subsequently, molecular dynamics simulation of the small molecule–protein complexes was carried out using GROMACS software with energy minimization and simulation times up to 100 ns, and the root mean square deviation (RMSD), root mean square fluctuation (RMSF), and binding free energy between the protein and the small molecule were calculated [25,26].

2.6. Statistical analysis

All statistical analyses were executed in R (version 4.3.0). The Wilcoxon test was employed as the primary method for assessing differences in gene expression of sequencing data. In MR studies, multiple comparisons were rigorously controlled using the Benjamini-Hochberg (BH) procedure. The MR effect size was expressed using odds ratio (OR) with a 95 % confidence interval (95 % CI).

3. Results

3.1. Screening unique cell populations and DEGs in scRNA-seq

First, we identified five major cell types from the PBMCs of CT, CD, and RA patients based on marker genes such as MS4A1, PPBP,
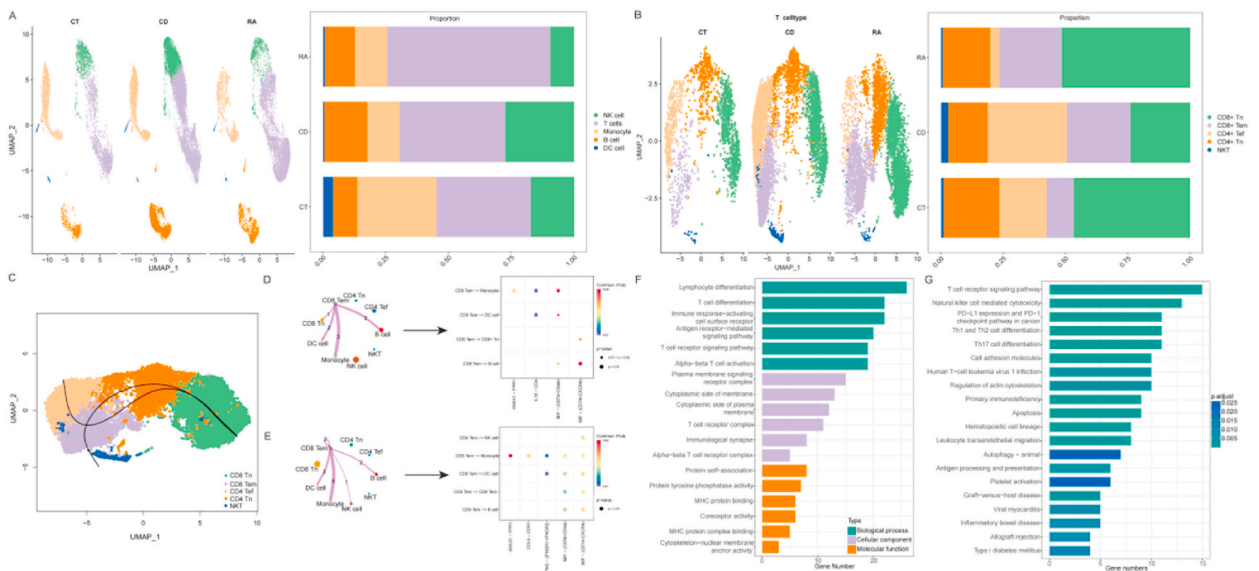

Fig. 2. scRNA-seq analysis in the PBMC of CT, RA and CD patients. (A) UMAP plot and proportion of distinct cell subsets in PBMC; (B) UMAP plot and proportion of various T cell subgroups; (C) Pseudotime analysis of T cells inferred by Slingshot; (D) Intercellular ligand–receptor prediction among CD8 Tem and other PBMC cells revealed by CellChat in CD patient; (E) Intercellular ligand–receptor pre-diction among CD8 Tem and other PBMC cells revealed by CellChat in RA patient; (F) GO enrichment analysis of specifically highly expressed genes in CD8 Tem; (G) KEGG pathway enrichment of specifically highly expressed genes in CD8 Tem.

FCER1A, GNLY, CD3E, and CD14: natural killer (NK) cells, T cells, monocytes, B cells, and dendritic cells (DCs). Among them, T cells were the most prevalent in PBMCs, and their proportion increased significantly in the disease groups (Fig. 2A). Therefore, we focused on T cells for further classification. According to previous research by Zhang Zemin's team, T cells can be categorized into CD4 naive T cells (Tn), CD4 effector T cells (Tef), CD8 Tn cells, CD8 effector memory T cells (Tem), and NK and T cells [27]. We found that the proportion of CD8 Tem cells was significantly increased in both diseases (Fig. 2B). The results of T cell lineage analysis and pseudotime inference are depicted in Fig. 2C, where all trajectories pass through CD8 Tem cells, suggesting an active proliferation phase of CD8 Tem cells during the development of both diseases. As shown in Fig. 2D, CD8 Tem cells interact with four other cell types via four signalling pathways, with macrophage migration inhibitory factor (MIF) exhibiting the strongest outward signalling in CD patients. Fig. 2E illustrates that CD8 Tem cells interact with five other cell types in RA patients through five signalling pathways, again with strong outward signalling via the MIF pathway. However, compared to those in CD patients, CD8 Tem cells in individuals with RA did not show an association with CD8 Tn cells but could regulate NK cells as well as their own population. We collected and organized a total of 138 DEGs in CD8 Tem cells (Supplementary Table 1). Subsequently, we conducted GO and KEGG pathway enrichment analyses of the DEGs. The GO enrichment analysis revealed significant involvement in various biological processes, including lymphocyte differentiation, T cells differentiation, and immune activation. Cellular components included the plasma membrane signalling receptor complex, cytoplasmic side of the membrane, and cytoplasmic side of the plasma membrane. The associated molecular functions included protein self-association, protein tyrosine phosphatase activity, and MHC protein binding, as shown in Fig. 2F. Furthermore, the KEGG pathway enrichment analysis indicated strong associations with pathways such as the T cell receptor pathway, NK cell receptor pathway, and MHC protein binding. Notably, the analysis revealed enrichment in the inflammatory bowel disease-related pathway, aligning with the purpose of our research (Fig. 2G).

3.2. Screening the DEGs for comorbidity-causing genes

The results of the genetic variation instrument variables for DEGs, variable harmonization, and Steiger filtering are presented in Supplementary Table 2, ensuring alignment between the exposure and outcome instrumental variables. After performing BH correction, a total of 12 gene-CD pairs were obtained (Fig. 3A, Fig. 4A and Supplementary Table 3). Of these, four demonstrated a protective effect on disease development, including annexin A1 (ANXA1, OR = 0.53; 95%CI, 0.43–0.66; P = 5.39E−07), granzyme H (GZMH, OR = 0.53; 95%CI, 0.41–0.68; P = 1.20E−05), myocardial infarction-associated transcript (MIAT, OR = 0.43; 95%CI, 0.31–0.58; P = 1.49E−06), and T cell immunoreceptor with Ig and ITIM domains (TIGIT, OR = 0.63; 95%CI, 0.48–0.82; P = 0.004). Conversely, the remaining eight exacerbate disease progression, including guanylate binding protein 5 (GBP5, OR = 1.43; 95%CI, 1.26–1.62; P = 5.39E−07), integrin subunit alpha 4 (ITGA4, OR = 1.14; 95%CI, 1.03–1.27; P = 0.046), mitochondrial ribosomal protein L10 (MRPL10, OR = 1.11; 95%CI, 1.03–1.19; P = 0.022), myb proto-oncogene like 1 (MYBL1, OR = 1.23; 95%CI, 1.05–1.45; P = 0.046), protein tyrosine phosphatase non-receptor type 4 (PTPN4, OR = 2.07; 95%CI, 1.58–2.71; P = 1.49E−06), sad1 and UNC84 domain containing 2 (SUN2, OR = 1.07; 95%CI, 1.02–1.13; P = 0.04), syne-2 (SYNE2, OR = 2.37; 95%CI, 1.72–3.26; P = 1.49E−06) and tumor necrosis factor alpha-induced protein 3 (TNFAIP3, OR = 1.41; 95%CI, 1.15–1.72; P = 0.06). Additionally, the P values of the MR analysis results for RA were also adjusted, and 5 causal genes remained (Figs. 3B and 4B and Supplementary Table 3), among which overexpression of FK506 binding protein 11 (FKBP11, OR = 5.01; 95%CI, 2.95, 8.5; P = 1.67E−07) and MYBL1 (OR = 1.45; 95%CI, 1.14–1.85; P = 0.04) increased the risk of RA, while on the contrary cathepsin W (CTSW, OR = 0.33; 95%CI, 0.21–0.52; P = 6.57E-05), granzyme B (GZMB, OR = 0.44; 95%CI, 0.29, 0.69; P = 0.005) and killer cell lectin-like receptor D1 (KLRD1, OR = 0.35; 95%CI, 0.22–0.57; P = 0.001) reduced the associated risk. Both MR results revealed that MYBL1 is a pathogenic gene shared between CD and RA. Meanwhile, no heterogeneity was detected for the genes analysed in the primary analysis (Supplementary Table 4).

Subsequently, since MYBL1 is a common causative gene for both CD and RA, bidirectional MR analysis was conducted for MYBL1
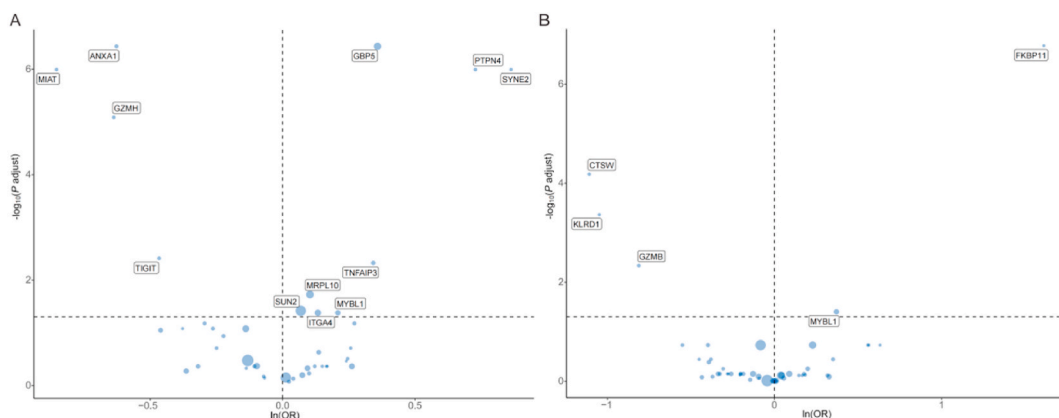

Fig. 3. MR results for DEGs and the risk of CD and RA. Volcano plots of the MR results for DEGs on the risk of CD (A) and RA (B), re-spectively. The size of the points on the graph is determined by the PVE value. Dashed horizontal black line corresponded to ad just P = 0.05 (BH). ln: natural logarithm; PVE: proportion of variance explained.
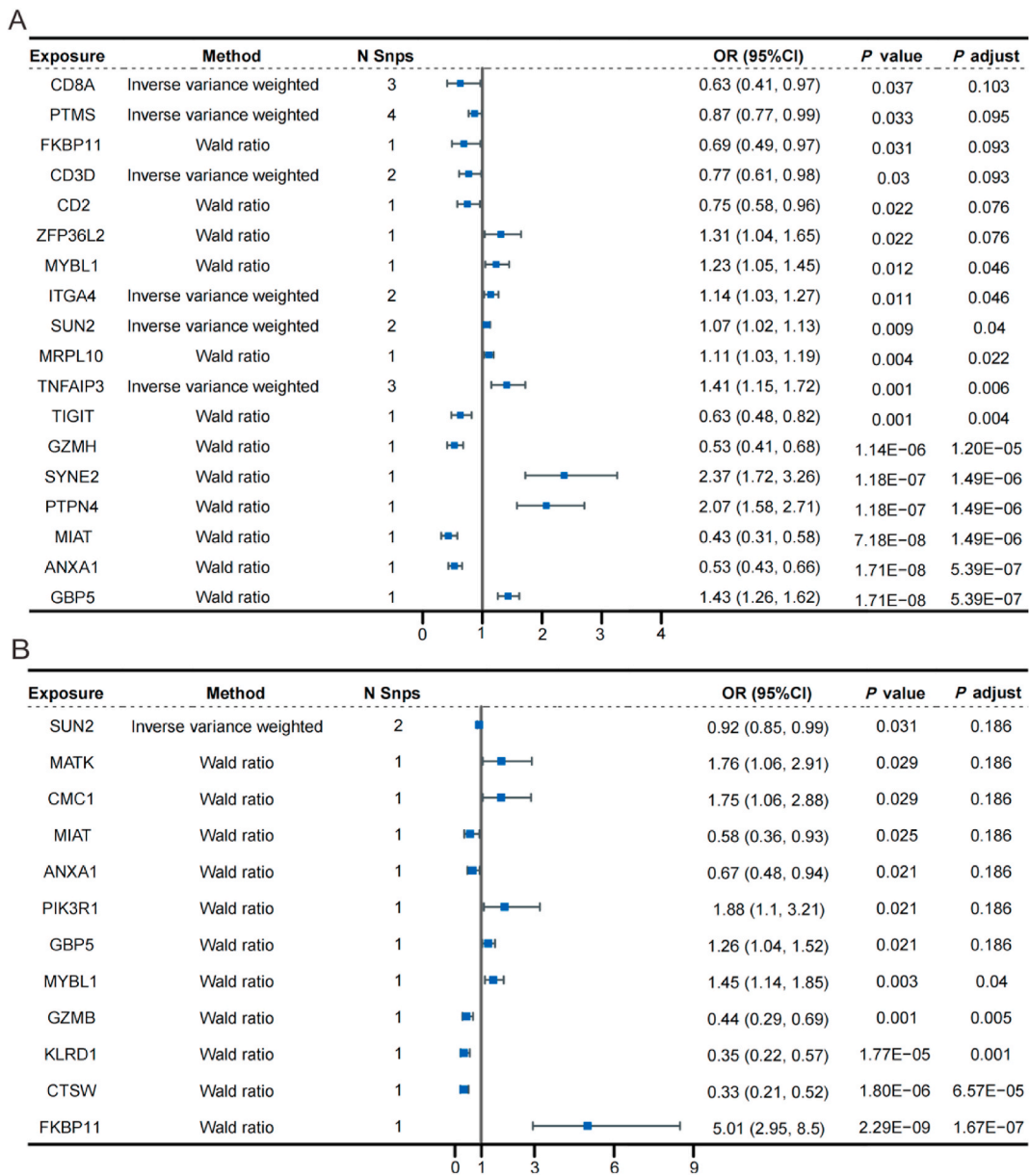
A

Exposure	Method	N Snps	OR (95%CI)	P value	P adjust
CD8A	Inverse variance weighted	3	0.63 (0.41, 0.97)	0.037	0.103
PTMS	Inverse variance weighted	4	0.87 (0.77, 0.99)	0.033	0.095
FKBP11	Wald ratio	1	0.69 (0.49, 0.97)	0.031	0.093
CD3D	Inverse variance weighted	2	0.77 (0.61, 0.98)	0.03	0.093
CD2	Wald ratio	1	0.75 (0.58, 0.96)	0.022	0.076
ZFP36L2	Wald ratio	1	1.31 (1.04, 1.65)	0.022	0.076
MYBL1	Wald ratio	1	1.23 (1.05, 1.45)	0.012	0.046
ITGA4	Inverse variance weighted	2	1.14 (1.03, 1.27)	0.011	0.046
SUN2	Inverse variance weighted	2	1.07 (1.02, 1.13)	0.009	0.04
MRPL10	Wald ratio	1	1.11 (1.03, 1.19)	0.004	0.022
TNFAIP3	Inverse variance weighted	3	1.41 (1.15, 1.72)	0.001	0.006
TIGIT	Wald ratio	1	0.63 (0.48, 0.82)	0.001	0.004
GZMH	Wald ratio	1	0.53 (0.41, 0.68)	1.14E-06	1.20E-05
SYNE2	Wald ratio	1	2.37 (1.72, 3.26)	1.18E-07	1.49E-06
PTPN4	Wald ratio	1	2.07 (1.58, 2.71)	1.18E-07	1.49E-06
MIAT	Wald ratio	1	0.43 (0.31, 0.58)	7.18E-08	1.49E-06
ANXA1	Wald ratio	1	0.53 (0.43, 0.66)	1.71E-08	5.39E-07
GBP5	Wald ratio	1	1.43 (1.26, 1.62)	1.71E-08	5.39E-07

B

Exposure	Method	N Snps	OR (95%CI)	P value	P adjust
SUN2	Inverse variance weighted	2	0.92 (0.85, 0.99)	0.031	0.186
MATK	Wald ratio	1	1.76 (1.06, 2.91)	0.029	0.186
CMC1	Wald ratio	1	1.75 (1.06, 2.88)	0.029	0.186
MIAT	Wald ratio	1	0.58 (0.36, 0.93)	0.025	0.186
ANXA1	Wald ratio	1	0.67 (0.48, 0.94)	0.021	0.186
PIK3R1	Wald ratio	1	1.88 (1.1, 3.21)	0.021	0.186
GBP5	Wald ratio	1	1.26 (1.04, 1.52)	0.021	0.186
MYBL1	Wald ratio	1	1.45 (1.14, 1.85)	0.003	0.04
GZMB	Wald ratio	1	0.44 (0.29, 0.69)	0.001	0.005
KLRD1	Wald ratio	1	0.35 (0.22, 0.57)	1.77E-05	0.001
CTSW	Wald ratio	1	0.33 (0.21, 0.52)	1.80E-06	6.57E-05
FKBP11	Wald ratio	1	5.01 (2.95, 8.5)	2.29E-09	1.67E-07

Fig. 4. Forest Plots of the results of the MR analysis for CD (A) and RA (B), respectively. The P values were adjusted by the BH method. SNP: single nucleotide polymorphism.

individually and in relation to comorbidities, revealing no evidence of a causal effect between disease and MYBL1 (Supplementary Table 5). However, phenotypic scanning identified a between MYBL1 (rs1060242) and daytime napping (Supplementary Table 6). Unexpectedly, however, the Bayesian colocalization analysis showed that MYBL1 did not share genetic variants with CD (PPH_4 = 0.199) and RA (PPH_4 = 0.384), as detailed in Fig. 5 and Supplementary Table 7.

3.3. Role of MYBL1 in CD8 Tem

As shown in Fig. 6A, MYBL1 gene expression in PBMCs occurs predominantly in CD8 Tem cells. During the maturation process of CD8 Tem cells, MYBL1 gradually becomes downregulated, leading us to categorize them into MYBL1+ and MYBL1- CD8 Tem sub-populations (Supplementary Table 8 and Supplementary Fig. 1). In CD patients, MYBL1- CD8 Tem cells exhibited stronger interactions with other cell types than did MYBL1+ CD8 Tem cells. MYBL1-cells achieved these interactions primarily through MIF, while MYBL1+ cells utilized IL16–CD4 (Fig. 6B). On the other hand, in RA patients, we observed similar intercellular interactions between the MYBL1+ and MYBL1-cell populations, both mediated by MIF. However, the MYBL1+ cell population was capable of interacting with
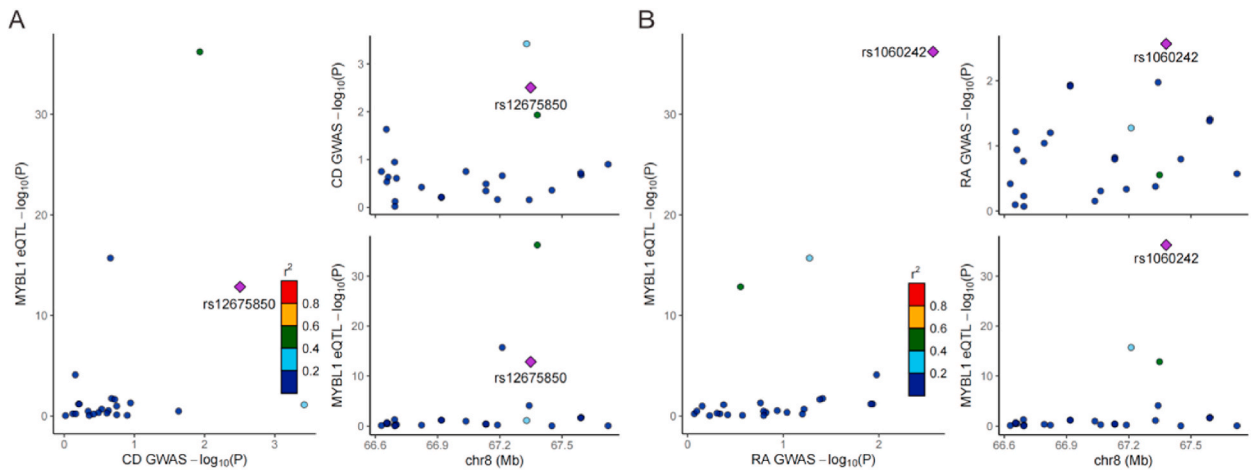
Fig. 5. Regional association plots of MYBL1 and CD (A) and RA (B), respectively. Diamond purple points represented the SNP that with the minimal sum of P value in corresponded gene GWAS and disease GWAS.

both the MYBL1+ and MYBL1-cell populations (Fig. 6C). In terms of cellular metabolic activities, the MYBL1+ cell population displayed heightened activity in drug metabolism involving other enzymes, whereas the MYBL1-cell population exhibited increased vigour in xenobiotic metabolism mediated by cytochrome P450 (Fig. 6D and E).

3.4. Molecular docking and dynamics simulatuons

Four potential drug treatments, namely, arsenic trioxide, azathioprine, cyclosporine, and tretinoin, were screened from the CTD database as candidate inhibitors of MYBL1 expression in both CD and RA (Supplementary Table 9). Subsequently, two of these small molecule medicines, azathioprine (PubChem CID: 2265, docking score: −5.5 kcal/mol) and Tretinoin (PubChem CID: 444795, binding energy: −6.3 kcal/mol), were selected for molecular docking with MYBL1 individually (Supplementary Fig. 2 and Fig. 7A). The results revealed that of these two drugs, tretinoin exhibited a stronger affinity towards MYBL1. Molecular dynamics simulations were then employed to validate the binding capability of tretinoin with MYBL1.

As shown in Fig. 7B, over the initial 100 ns of the simulation, the RMSD value between the Tretinoin/MYBL1 complexes gradually stabilized within a narrow range, indicating a more stable binding interaction. The RMSF values, shown in Fig. 7C illustrates the flexibility of each residue within the molecule. Moreover, the MM/GBSA analysis yielded an average binding free energy value of −25.56 ± 2.68 kcal/mol, suggesting that tretinoin can form a stable binding complex with MYBL1, thereby exerting pharmacological effects.

4. Discussion

To the best of our knowledge, this is the first study to explore the comorbidity of CD and RA by integrating scRNA-seq data from PBMCs with eQTL, MR, and colocalization analyses. Importantly, we identified MYBL1 as a critical pathogenic target, which is highly expressed in CD8 Tem cells. To substantiate our findings, we employed molecular docking, and molecular dynamics simulation, which further corroborated the reliability of the identified potential targets and therapeutic agents. This comprehensive and integrative approach not only enhances our understanding of the shared pathogenesis of CD and RA but also paves the way for the development of novel therapeutic strategies for these debilitating diseases.

T cells, a critical component of PBMCs, play an essential role in maintaining immune homeostasis and combating various diseases. Notably, the dysfunction of T cells is associated with numerous diseases, including autoimmune disorders, infectious diseases, and cancer, underscoring their crucial role in immune balance and disease development [28]. Our study aligns with previous research indicating an elevated trend of CD8 Tem in various diseases, underscoring their pivotal role in adaptive responses and long-term immunity [29]. Moreover, we found that both CD4 Tn and CD8 Tn cells can be converted to CD8 Tem cells through a series of regulatory mechanisms involving antigen expression, activation, differentiation, contraction, and survival. Interestingly, our study highlights the significant role of the MIF pathway in the interaction between CD8 Tem cells and multiple other immune cell types, which expands previous research finding that CD8 Tem cells can interact with multiple immune cell types [30]. GO and KEGG enrichment analyses of DEGs specifically overexpressed in CD8 Tem cells showed that these DEGs were mainly involved in pathways related to lymphocyte cell differentiation, regulation, migration, apoptosis, autophagy, and inflammatory diseases, which is consistent with their biological significance.

Future research on CD8 Tem cells could pave the way for novel diagnostic and therapeutic strategies targeting these cells, potentially expanding treatment options in clinical settings. In addition, DEGs in CD8 Tem cells are expected to be new therapeutic targets, so we combined MR and colocalization approaches to comprehensively analyse the causal proteins of CD and RA comorbidity
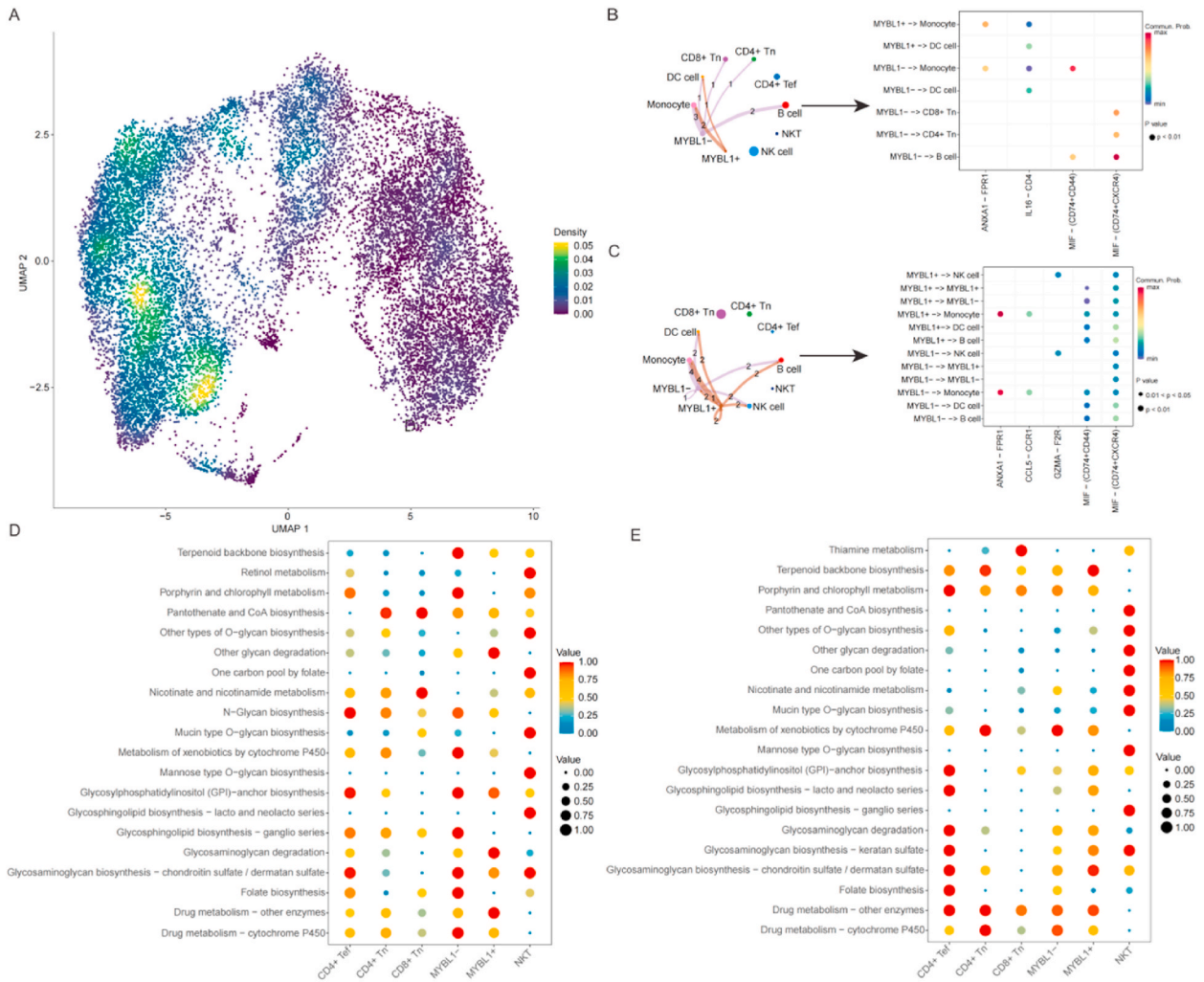
Fig. 6. The role of MYBL1 in T cells. (A) Nebulosa expression density for MYBL1 in T cells. (B) Intercellular ligand–receptor prediction among MYBL1(±) CD8 Tem and other PBMC cells revealed by CellChat in CD patient. (C) Intercellular ligand–receptor prediction among MYBL1(±) CD8 Tem and other PBMC cells revealed by CellChat in RA patient. (D)The metabolic activity of PBMC in CD patients. (E)The metabolic activity of PBMC in RA patients.

towards a clinical translation of previous GWAS findings. To avoid genetic confounding due to reverse causality and linkage disequilibrium inherent in MR, we employed Steiger directionality filtering and reverse MR analysis. We identified MYBL1 as a causal gene in the comorbidity of CD and RA, with colocalization results supporting this genetic variation as shared by both disorders. Phenotypic scanning revealed an association between an MYBL1-related SNP (rs1060242) and daytime napping, a seemingly unrelated trait.

However, recent prospective cohort studies have linked daytime napping with inflammatory bowel disease [31]. Furthermore, daytime napping has been associated with pain, depression, anxiety, fatigue, memory difficulties, and insomnia, all of which are relevant to CD and RA [32,33]. MYBL1, a member of the MYB transcription factor family, plays crucial roles in cell cycle progression, differentiation, and apoptosis. Its aberrant expression is closely associated with human cancers, and it has been proposed as a novel therapeutic target in renal cancer through remodelling of the tumour microenvironment [34]. However, the role of MYBL1 in autoimmune diseases has been largely overlooked.

We further investigated the role of MYBL1 in T cells. Utilizing a gene-weighted 2D kernel density plot, we observed a distinct overexpression of MYBL1 in CD8 Tem cells, with its regulation in these cells negatively correlated with developmental time. Based on the regulatory status of MYBL1, we stratified CD8 Tem cells into MYBL1+ and MYBL1-subgroups. We hypothesized that MYBL1+ CD8 Tem cells could be the pathogenic cell subgroup involved in this comorbidity. Cell interaction analysis revealed reduced interactions between MYBL1+ CD8 Tem cells and other immune cells, and distinct metabolic activities were observed between the MYBL1+ and MYBL1-subgroups. Subsequent whole-blood transcriptomic analysis showed a significant overexpression of MYBL1 in the disease group, indirectly supporting our hypothesis. However, it is worth noting that although we identified MYBL1 as a potential therapeutic
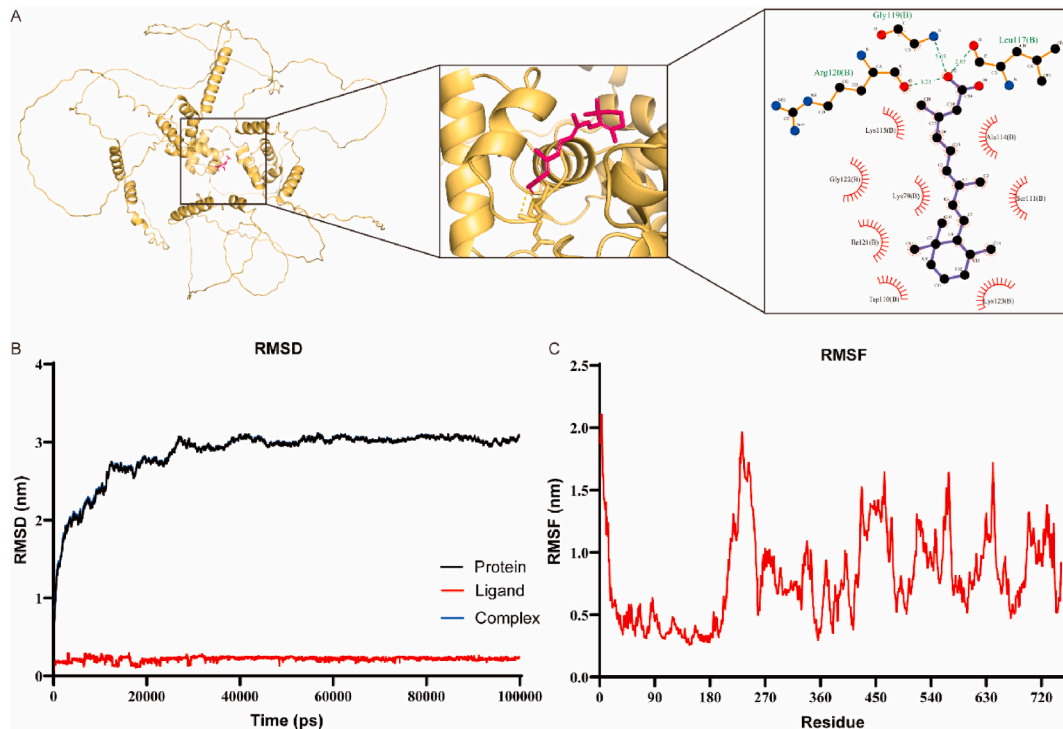
Fig. 7. Molecular docking (A) and dynamics simulation between tretinoin and MYBL1. RMSD (B) of the tretinoin and MYBL1. RMSF (C) of the MYBL1 in the complex. RMSD: root mean square deviation; RMSAF: root mean square fluctuation.

target for comorbidity, its role in the development or maintenance of these disorders requires further interpretation.

After identifying this potential target, we mined four potentially valuable medicines from the CTD database. Based on the compounds' biological activity, safety, reuse potential, and molecular binding energy, we selected tretinoin for further analysis. Tretinoin, a derivative of vitamin A, is clinically used for the treatment of acne and acute promyelocytic leukemia [35]. However, emerging evidence suggests that it may have broader potential applications. Studies have indicated that tretinoin can modulate T cell function and differentiation, playing a potential role in various diseases [36]. However, its precise mechanisms of action remain unclear. In our study, we attempted to elucidate these mechanisms using molecular docking and dynamic simulations. We propose that tretinoin may exert its therapeutic effects in comorbid conditions by inhibiting the expression of MYBL1, thereby regulating the function of CD8 Tem cells. Future work will involve biological experiments to further validate these findings. Additionally, we plan to rationally modify the structure of tretinoin or design suitable drug carriers to enhance its efficacy and minimize side effects.

5. Limitations

This study had some limitations that should be acknowledged. First, our PBMC scRNA-seq data were obtained from public databases, which were of limited sample size and lacked data from patients codiagnosed with CD and RA. Second, both the eQTL and GWAS data were based on studies conducted in populations of European ancestry, limiting the generalizability of our findings to other ethnic groups. Further research in non-European populations is warranted to facilitate the clinical translation of these findings. Last, although our results suggest that tretinoin could be a potential therapeutic agent for comorbidity, the molecular docking and molecular dynamics simulation results are suggestive rather than conclusive. Further in-depth studies are needed to validate these preliminary findings. In the future, conducting additional animal studies and preclinical trials based on the findings of this study will aid in further elucidating the pathogenesis and therapeutics of RA and CD comorbidity.

6. Conclusions

In conclusion, our comprehensive analysis indicates that CD8 Tem cells play a role in the pathogenesis of comorbid CD and RA. Specifically, the MYBL1 gene is causally associated with the risk of comorbid CD and RA and represents a potential drug target. Moreover, tretinoin shows promise as a potential treatment for comorbid CD and RA. This study contributes to laying the groundwork for a more profound understanding of comorbid CD and RA by offering a unique and comprehensive perspective into the sequencing data of PBMCs, as well as potential therapeutic avenues.

Ethical statement

Review and/or approval by an ethics committee was not needed for this study because all the data utilized in this study were obtained from previously published research and publicly accessible databases.

Funding

Data availability statement

GWAS data can be downloaded from the GWAS Catalog using the accession numbers ebi-a-GCST003044 and ebi-a-GCST005569. Sequencing data can be accessed from the GEO database using GSM4976997, GSM4977001, GSM4976999, GSM4977003, and GSE159117.

CRediT authorship contribution statement

Qiu Dong: Writing – original draft, Data curation, Conceptualization. **Xiaoting Wu:** Investigation, Funding acquisition, Formal analysis. **Tsz Mok:** Project administration, Methodology. **Gaohan Cai:** Software, Resources. **Zhengang Zha:** Supervision, Software. **Guorong She:** Visualization, Validation. **Junyuan Chen:** Writing – review & editing, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e32406.

References

[1] K. Cushing, P.D.R. Higgins, Management of crohn disease, JAMA 325 (1) (2021) 69–80.

[2] D. Aletaha, J.S. Smolen, Diagnosis and Management of rheumatoid arthritis: a review, JAMA 320 (13) (2018) 1360–1372.

[3] J. Park, G.H. Jeong, M. Song, D.K. Yon, S.W. Lee, et al., The global, regional, and national burden of inflammatory bowel diseases, 1990–2019: a systematic analysis for the global burden of disease study 2019, Dig. Liver Dis. 55 (10) (2023) 1352–1359.

[4] P. Gao, X. Pan, S. Wang, S. Guo, Z. Dong, et al., Identification of the transcriptome signatures and immune-inflammatory responses in postmenopausal osteoporosis, Heliyon 10 (1) (2024) e23675, https://doi.org/10.1016/j.heliyon.2023.e23675.

[5] Y. Chen, L. Chen, C. Xing, G. Deng, F. Zeng, et al., The risk of rheumatoid arthritis among patients with inflammatory bowel disease: a systematic review and meta-analysis, BMC Gastroenterol. 20 (1) (2020) 192.

[6] C. Meisinger, D. Freuer, Rheumatoid arthritis and inflammatory bowel disease: a bidirectional two-sample Mendelian randomization study, Semin. Arthritis Rheum. 55 (2022) 151992.

[7] P.B. Allen, P. Olivera, P. Emery, D. Moulin, J.Y. Jouzeau, et al., Review article: moving towards common therapeutic goals in Crohn's disease and rheumatoid arthritis, Aliment. Pharmacol. Ther. 45 (8) (2017) 1058–1072.

[8] D.Y. Zheng, Y.N. Wang, Y.H. Huang, M. Jiang, C. Dai, Effectiveness and safety of upadacitinib for inflammatory bowel disease: a systematic review and meta-analysis of RCT and real-world observational studies, Int. Immunopharm. 126 (2024) 111229, https://doi.org/10.1016/j.intimp.2023.111229.

[9] A. O'Toole, M. Lucci, J. Korzenik, Inflammatory bowel disease provoked by etanercept: report of 443 possible cases combined from an IBD referral center and the FDA, Dig. Dis. Sci. 61 (6) (2016) 1772–1774.

[10] D. Pu, L. Liu, N. Wang, D. Wang, Z. Zhang, B. Feng, Case report: single-cell mapping of peripheral blood mononuclear cells from a patient with both Crohn's disease and isolated congenital asplenia, Front. Immunol. 13 (2022) 959281.

[11] X. Song, Y. Zhang, L. Zhao, J. Fan, T. Peng, et al., Analyzation of the peripheral blood mononuclear cells atlas and cell communication of rheumatoid arthritis patients based on single-cell RNA-seq, Journal of Immunology Research 2023 (2023) 6300633.

[12] J. Lin, J. Zhou, Y. Xu, Potential drug targets for multiple sclerosis identified through Mendelian randomization analysis, Brain 146 (8) (2023) 3364–3372.

[13] U. Võsa, A. Claringbould, H.J. Westra, M.J. Bonder, P. Deelen, et al., Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression, Nat. Genet. 53 (9) (2021) 1300–1310, https://doi.org/10.1038/s41588-021-00913-z.

[14] C.N. Foley, J.R. Staley, P.G. Breen, B.B. Sun, P.D.W. Kirk, et al., A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits, Nat. Commun. 12 (1) (2021) 764.

[15] Y. Hao, T. Stuart, M. Kowalski, S. Choudhary, P. Hoffman, et al., Dictionary learning for integrative, multimodal, and scalable single-cell analysis, bioRxiv (2022).

[16] U. Võsa, A. Claringbould, H.-J. Westra, M.J. Bonder, P. Deelen, et al., Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression, Nat. Genet. 53 (9) (2021) 1300–1310.

[17] J.Z. Liu, S. van Sommeren, H. Huang, S.C. Ng, R. Alberts, et al., Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations, Nat. Genet. 47 (9) (2015) 979–986.

[18] S. Eyre, J. Bowes, D. Diogo, A. Lee, A. Barton, et al., High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis, Nat. Genet. 44 (12) (2012) 1336–1340.

[19] K. Street, D. Risso, R.B. Fletcher, D. Das, J. Ngai, et al., Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics, BMC Genom. 19 (1) (2018) 477.

[20] S. Jin, C.F. Guerrero-Juarez, L. Zhang, I. Chang, R. Ramos, et al., Inference and analysis of cell-cell communication using CellChat, Nat. Commun. 12 (1) (2021) 1088.

[21] G. Yu, L.-G. Wang, Y. Han, Q.-Y. He, clusterProfiler: an R Package for comparing biological themes among gene clusters, OMICS A J. Integr. Biol. 16 (5) (2012) 284–287.

[22] A.P. Davis, T.C. Wiegers, R.J. Johnson, D. Sciaky, J. Wiegers, C.J. Mattingly, Comparative toxicogenomics database (CTD): update 2023, Nucleic Acids Res. 51 (D1) (2023) D1257–D1262.

[23] J. Eberhardt, D. Santos-Martins, A.F. Tillack, S. Forli, AutoDock vina 1.2.0: new docking methods, expanded force field, and Python bindings, J. Chem. Inf. Model. 61 (8) (2021) 3891–3898.

[24] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, et al., Highly accurate protein structure prediction with AlphaFold, Nature 596 (7873) (2021) 583–589.

[25] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, et al., GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX 1–2 (2015) 19–25.

[26] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, GROMACS: fast, flexible, and free, J. Comput. Chem. 26 (16) (2005) 1701–1718.

[27] L. Zhang, X. Yu, L. Zheng, Y. Zhang, Y. Li, et al., Lineage tracking reveals dynamic relationships of T cells in colorectal cancer, Nature 564 (7735) (2018) 268–272.

[28] U.H. von Andrian, C.R. Mackay, T-cell function and migration. Two sides of the same coin, N. Engl. J. Med. 343 (14) (2000) 1020–1034.

[29] P. Khandelwal, A. Lane, V. Chaturvedi, E. Owsley, S.M. Davies, et al., Peripheral blood CD38 bright CD8+ effector memory T cells predict acute graft-versus-host disease, Biol. Blood Marrow Transplant.: Journal of the American Society for Blood and Marrow Transplantation 21 (7) (2015) 1215–1222.

[30] H.G. Duong, E.J. Choi, P. Hsu, N.R. Chiang, S.A. Patel, et al., Identification of CD8 + T-cell-immune cell communications in ileal crohn's disease, Clin. Transl. Gastroenterol. 14 (5) (2023) e00576.

[31] S. Yuan, Y. Sun, X. Tan, J. Geng, J. Sun, et al., Sleep duration and daytime napping in relation to incident inflammatory bowel disease: a prospective cohort study, Aliment. Pharmacol. Ther. 57 (5) (2023) 475–485.

[32] A. Theadom, M. Cropley, T. Kantermann, Daytime napping associated with increased symptom severity in fibromyalgia syndrome, BMC Muscoskel. Disord. 16 (2015) 13.

[33] C. Alexandre, A. Latremoliere, A. Ferreira, G. Miracca, M. Yamamoto, et al., Decreased alertness due to sleep loss increases pain sensitivity in mice, Nat. Med. 23 (6) (2017) 768–774.

[34] Y. Cicirò, A. Sala, MYB oncoproteins: emerging players and potential therapeutic targets in human cancer, Oncogenesis 10 (2) (2021) 1–15.

[35] N. Mosquera, A. Rodriguez-Trillo, F.J. Blanco, A. Mera-Varela, A. Gonzalez, C. Conde, All-Trans retinoic acid inhibits migration and invasiveness of rheumatoid fibroblast-like synoviocytes, J. Pharmacol. Exp. Therapeut. 372 (2) (2020) 185–192.

[36] J.A. Hall, J.R. Grainger, S.P. Spencer, Y. Belkaid, The role of retinoic acid in tolerance and immunity, Immunity 35 (1) (2011) 13–22.