RESEARCH

Large-scale integrative analysis of juvenile idiopathic arthritis for new insight into its pathogenesis

Daeun Kim[1], Jaeseung Song[1], Nicholas Mancuso[2,3], Serghei Mangul[3,4], Junghyun Jung[1,2,5*†] and Wonhee Jang[1*†]

Abstract

Background   Juvenile idiopathic arthritis (JIA) is one of the most prevalent rheumatic disorders in children and is classified as an autoimmune disease (AID). While a robust genetic contribution to JIA etiology has been established, the exact pathogenesis remains unclear.

Methods   To prioritize biologically interpretable susceptibility genes and proteins for JIA, we conducted transcriptome-wide and proteome-wide association studies (TWAS/PWAS). Then, to understand the genetic architecture of JIA, we systematically analyzed single-nucleotide polymorphism (SNP)-based heritability, a signature of natural selection, and polygenicity. Next, we conducted HLA typing using multi-ethnicity RNA sequencing data. Additionally, we examined the T cell receptor (TCR) repertoire at a single-cell level to explore the potential links between immunity and JIA risk.

Results   We have identified 19 TWAS genes and two PWAS proteins associated with JIA risks. Furthermore, we observe that the heritability and cell type enrichment analysis of JIA are enriched in T lymphocytes and HLA regions and that JIA shows higher polygenicity compared to other AIDs. In multi-ancestry HLA typing, B*45:01 is more prevalent in African JIA patients than in European JIA patients, whereas DQA1*01:01, DQA1*03:01, and DRB1*04:01 exhibit a higher frequency in European JIA patients. Using single-cell immune repertoire analysis, we identify clonally expanded T cell subpopulations in JIA patients, including *CXCL13*⁺*BHLHE40*⁺ T$_H$ cells which are significantly associated with JIA risks.

Conclusion   Our findings shed new light on the pathogenesis of JIA and provide a strong foundation for future mechanistic studies aimed at uncovering the molecular drivers of JIA.

Keywords   Juvenile idiopathic arthritis, Transcriptome-wide and proteome-wide association studies, T cell receptor (TCR) repertoire, Multi-ethnicity RNA typing

---

†Junghyun Jung and Wonhee Jang jointly supervised this work.

*Correspondence:
Junghyun Jung
junghyun.jung@cshs.org
Wonhee Jang
wany@dongguk.edu
Full list of author information is available at the end of the article

Kim *et al. Arthritis Research & Therapy*   (2024) 26:47

Page 2 of 19

Introduction

Juvenile idiopathic arthritis (JIA) is one of the most common rheumatic diseases in children; it is mainly regarded as an autoimmune disease (AID) whose clinical manifestations include persistent limping, painful joints, stiffness, and inflammation [1, 2]. The exact JIA pathogenesis remains unclear; however, it has been demonstrated that there is a strong genetic contribution to the etiology of JIA. The single-nucleotide polymorphism (SNP)-based heritability for JIA was estimated to be 73%, among the most highly heritable pediatric AIDs. Even though genome-wide association studies (GWASs) have identified many risk variants of JIA, most are located in non-coding regions, making it difficult to interpret their functional consequences [3]. By integrating GWASs with expression quantitative trait loci (eQTL), transcriptome-wide association studies (TWASs) provide a powerful approach to prioritize susceptibility genes affected by risk variants [4].

Investigating the genetic architecture of complex diseases is essential for understanding the genetic basis of phenotypic variations and evolutions [5]. For complex diseases, natural selection plays an essential role in forming the genetic architecture and provides valuable insights into biological mechanisms [6]. In AIDs, the signature of negative selection is significant because it sheds light on how the human immune system has evolved to defend against pathogens while avoiding harmful responses against self-tissues.

Moreover, given that JIA is considered an AID, it is also important to further understand the underlying mechanisms of immune responses affecting the JIA etiologies. Human leukocyte antigen (HLA) molecules presenting peptide antigens to receptors on T lymphocytes are highly polymorphic at their peptide-binding site and mediate the adaptive immune responses [5]. Moreover, T cell receptors (TCRs) interacting with HLA molecules are essential components of adaptive immune responses [7]. Through the somatic recombination of TCRs, self-reactive T cells can be produced, and the recognition of self-antigens can affect the development of AIDs [8]. Even though there is some evidence that T cells are involved in JIA etiologies, their contributions to the pathogenesis of AIDs including JIA have not been completely revealed [9, 10].

Herein, we prioritized susceptibility genes/proteins for JIA by conducting a multi-tissue TWAS and proteome-wide association study (PWAS) by integrating JIA GWAS summary statistics data (n = 3,305 cases and n = 9,196 controls) with reference eQTL/protein QTL (pQTL) panels (n = 14,037). We identified 19 genes and two proteins associated with JIA risk using TWAS and PWAS. We then estimated disease heritability and signatures of natural selection to understand the genetic architecture and to prioritize the most relevant tissue and cell types of JIA. We observed that the heritability of JIA was enriched in T lymphocytes and HLA regions and that JIA showed higher polygenicity than other AIDs. Next, HLA typing was conducted using multi-ancestry RNA sequencing (RNA-seq) data, and TCR repertoire analysis was performed at a single-cell level to investigate the associations between immunity and JIA risks. We found some HLA types, such as B*45:01, DQA1*01:01, DQA1*03:01, and DRB1*04:01, which were more frequent in the European or African JIA patients. In addition, we identified clonally expanded T cell subpopulations in JIA patients, among which $CXCL13^+BHLHE40^+$ T cells were significantly associated with JIA risks at the single-cell level. We believe that these findings could provide new insights into understanding the underlying mechanisms of JIA etiology.

Methods

Genome-wide association summary statistics of JIA

GWAS summary statistics data of JIA were retrieved from the GWAS catalog and the most recent dataset was used for this study (catalog ID: GCST90010715) [11, 12]. Details on the process of genotyping and quality control were described by López-Isac et al. [12]. The summary statistics of JIA were computed only for the European population (n = 3305 JIA patients and n = 9196 control subjects), and the JIA samples consist of 8 subtypes (n = 860 persistent oligoarthritis, oligo-JIA; n = 505 extended oligo-JIA; n = 825 rheumatoid factor-negative polyarthritis, RF-negative poly-JIA; n = 195 RF-positive poly-JIA; n = 205 systemic JIA, sJIA; n = 252 enthesitis-related arthritis, ERA; n = 225 juvenile psoriatic arthritis, JPsA; n = 238 Undifferentiated/Missing). The GWAS study resulted in 6,334,221 SNPs with minor allele frequency (MAF) $\geq$ 1%. The JIA summary statistics file was converted into a sumstats-formatted file by LD score (LDSC) software (v1.0.1) [13].

Transcriptome-wide association study and proteome-wide association study

To identify susceptibility genes associated with the pathogenesis of JIA, a TWAS was performed using functional summary-based imputation (FUSION) [4]. Briefly, TWAS identifies risk genes associated with the target disease by integrating GWAS summary statistics data with the reference eQTL data of specific tissues, considering linkage disequilibrium (LD) structures. Using the eQTL panels and LD information, the *cis*-genetic components of gene expression are imputed from the JIA summary statistics data. Then, the predicted gene expression is used for association tests with JIA risks to identify

significant associations between the gene expression and the disease. Twelve connective tissue panels from the Genotype-Tissue Expression project v7 (GTEx v7; $n = 449$), the Metabolic Syndrome in Men study (METSIM; $n = 563$), the Netherlands Twin Registry (NTR; $n = 1247$), and the Young Finns Study (YFS; $n = 1264$) were selected as expression weights for transcriptomic imputation (Table S1) [14–19]. For each panel, predictive models for gene expression were trained using *cis*-regulated genes by SNPs within ±500kb of the transcription start site and are significant for heritability (*cis-h^2*) with $P < 0.01$. The European LD information from the 1000 Genomes project accounted for the LD regions [20]. Due to the complex LD patterns, we excluded the TWAS associations from major histocompatibility complex (MHC) regions [21]. The significance threshold of TWAS associations was corrected with the Bonferroni correction ($P_{TWAS} < 7.55 \times 10^{-07}$, 0.05/66,196).

We conducted a PWAS using the pQTL data from the INTERVAL ($n = 3301$) [22] and Atherosclerosis Risk in Communities (ARIC; $n = 7213$) study [23] of European individuals. For each pQTL panel, predictive models were trained using *cis*-regulatory proteins with SNPs within ±500 kb of the same transcriptional start site, and the significance levels of *cis-h*2 in INTERVAL ($n = 1031$ models) and ARIC ($n = 1309$ models) were 0.05 and 0.01, respectively. The predictive models in the INTERVAL study were fitted using the sum of single effects (SuSiE) [24], and the elastic net was used in the ARIC study. As with TWAS, the PWAS associations in the MHC region were excluded. The same significance thresholds with the Bonferroni correction were used for the PWAS associations as the TWAS associations ($P_{PWAS} < 2.16 \times 10^{-05}$, 0.05/2311).

Fine-mapping of TWAS and PWAS associations
We performed fine-mapping of causal gene sets (FOCUS) to identify TWAS/PWAS associations responsible for the disease, estimating gene-trait associations at the GWAS risk regions while considering LD structures and controlling for pleiotropic SNP effects [25]. FOCUS calculates the posterior inclusion probability (PIP) per the TWAS/PWAS association and suggests credible gene sets containing susceptibility genes at a 90% confidence level. We applied FOCUS to the specific loci for each panel where the significant TWAS/PWAS associations ($P_{TWAS} < 7.55 \times 10^{-07}$ and $P_{PWAS} < 2.16 \times 10^{-05}$) identified by FUSION were detected. The weight database for FOCUS was generated from the GTEx and INTERVAL FUSION weights.

Pathway enrichment analysis
We conducted a TWAS-based Gene Set Enrichment Analysis (TWAS-GSEA) using the TWAS results from individual tissue panels [26]. The TWAS associations with a panel and the 12 eQTL panels containing information on the position of genes were used as inputs for this analysis. We retrieved curated gene sets from canonical pathways (CP), WikiPathways, Kyoto Encyclopedia of Genes and Genomes (KEGG), BioCarta, Reactome, and Pathway Interaction Database (PID) from the Molecular Signatures Database (MSigDB v7.2) [27–32].

A GWAS-based pathway enrichment analysis was carried out using the multi-marker analysis of genomic annotation (MAGMA v1.07) [33]. The SNPs were annotated to the corresponding genes, based on dbSNP v151 SNP locations for the European group and NCBI Build 37 gene definitions [20, 34]. Gene sets retrieved from the CPs of MSigDB were used [27–32].

Genetic correlation analyses with other traits
To estimate genetic correlations at the genome-wide level between JIA and other AID-like ($n = 10$) and non-AID-like ($n = 15$) traits, an LD score regression was performed using the GWAS summary statistics data of JIA and 25 traits using LDSC [13, 35]. Sumstats-formatted publicly available summary statistics (PASS) data of 22 traits, excluding type 1 diabetes (T1D), were retrieved from LD Hub [36]. The summary statistics data of T1D were obtained from the GWAS catalog (catalog ID: GCST005536) because T1D PASS data were not reported in the LD hub, although the comorbidity of JIA and T1D has previously been reported [11, 37, 38]. Summary statistics data of AID_{ALL} and AID_{SURE} traits, the remaining two traits, reported in UK Biobank (UKBB) were also used. The LD score data for the European samples were used for this analysis [20].

To estimate transcriptome-wide genetic correlations between JIA and the 23 traits, the TWAS results of these traits were retrieved from the TWAS-hub [4]. Eighteen of the 23 traits were obtained from the TWAS-hub and TWAS analyses of T1D, AID_{ALL}, and AID_{SURE} traits were conducted using the same procedure as JIA. Transcriptome-wide genetic correlations between JIA, T1D, AID_{ALL}, AID_{SURE}, and the 18 traits were estimated using the RHOGE [39].

Weighted gene co-expression network analysis
This study used an Affymetrix microarray dataset (GEO study: GSE13501 [40]) and two RNA-seq datasets (GEO study: GSE112057 [41] and GSE79970 [42]) containing mRNA expression data on JIA patients and healthy controls. The preprocessing and normalization of datasets were performed in accordance with Kim et al. [2]. The top 7000 most-expressed genes of the GSE13501 [40] dataset were selected for simplicity after normalization following Jung et al. [43]. A signed weighted gene co-expression

Kim *et al. Arthritis Research & Therapy*   (2024) 26:47

Page 4 of 19

network analysis (WGCNA) was performed to identify co-expression modules comprising positively correlated genes based on bi-weight mid-correlation [44]. A soft-thresholding power (β) of seven was selected for the network construction (scale-free $r^2 = 0.8$). The expression profile of each module was summarized by the module eigengene (ME). The minimum size of modules was 50 genes and paired modules with high ME ($r > 85$) were merged. With the co-expression modules, module preservation analyses were conducted using the microarray dataset as a reference set and each RNA-seq dataset as a test set with the co-expression modules [45]. The analyses were permuted up to 1000 times and Z-summary scores were computed to identify the preserved modules. To examine whether TWAS associations were enriched in the co-expression modules, GSEA was performed with the fgsea R package using the co-expression modules as the reference gene sets [46]. The TWAS associations from each panel, arranged by their Z-scores in descending order, were used as the pre-ranked gene sets. Functional annotation of the co-expression modules, in which TWAS associations were enriched, was carried out using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) [47].

JIA-relevant-tissue and cell-type analyses

LD score regression to specifically expressed genes (LDSC-SEG) v1.0.1 was applied to determine disease-relevant tissues and cell types in JIA [48]. Two types of precomputed expression datasets were downloaded for the analysis: (1) human RNA-seq data of 53 tissues/cell types from the GTEx [49] and human/mouse/rat array data of 152 tissues/cell types from Franke lab [50, 51] and (2) mouse array data of 292 immune cell types from ImmGen [52]. Two types of precomputed chromatin datasets were also downloaded: (1) human 431 tissue-specific epigenomic annotations from peaks for six epigenetic marks from Roadmap Epigenomics [53] and ENCODE projects [54] and (2) human ATAC-seq peaks from 13 cell types for human hematopoietic hierarchy [55]. Fetal data was excluded from the epigenomic data. The Bonferroni correction was applied to determine the significance levels.

Estimating the SNP-based heritability and expression-mediated heritability

We estimated SNP-based SNP heritability (h^2_{HESS}) in a set of 1702 independently partitioned genomic blocks [56] across the genome using Heritability Estimation from Summary Statistics (HESS) [39] v0.5.3-beta. To account for the large number of hypotheses tested, the Bonferroni correction was performed at $\alpha = 0.05/1702$ to determine significant levels. The mediated expression score

regression (MESC) [57] was used to estimate the proportion of heritability mediated by a *cis*-genetic component of gene expression levels (h^2_{med}/h^2_g). The MESC software was downloaded, along with its precomputed expression scores from the GTEx consortium and eQTLGen [58].

Genetic architecture analysis

To identify signatures of negative selection for JIA, we utilized summary-data-based BayesS (SBayesS) [59] in the GCTB software by estimating the joint posterior distribution of effect size and MAF. Based on the Markov-chain Monte Carlo sample, the posterior mean was used as a point estimation, and the posterior standard error was approximated by the standard deviation. Moreover, a sparse LD matrix was used for computational efficiency [59].

Estimation of HLA gene expression and HLA typing analysis

Consensus HLA typing analysis was carried out using the GSE112057 [41] RNA-seq dataset by seven HLA-typing software: seq2HLA (v2.3), arcasHLA (v0.2.0), HLAforest, HLA-VBSeq (v2), OptiType (v1.3.3), PHLAT (v1.0), and HLA typing from the high-quality dictionary (HLA-HD) (v1.2.1) [42, 60–66]. The GSE112057 [41] dataset contains expression data on 115 JIA patients (43 oligo-JIA, 46 poly-JIA, and 26 sJIA; 34 African JIA and 81 European JIA) and 12 control subjects. The reads mapped to chromosome 6 were detected by STAR software [67] (v2.5.3a) using the human reference genome (hg19) and were then analyzed by seven HLA-typing software with default settings. The International Immunogenetics Project/HLA database [68, 69] (v3.10.0) was used for arcasHLA, HLA-VBSeq, and HLA-HD. The consensus frequencies of HLA allele types (HLA class I: A, B, and C; HLA class II: DP alpha 1 [DPA1], DP beta 1 [DPB1], DQ alpha 1 [DQA1], DQ beta 1 [DQB1], DR alpha [DRA], DR beta 1 [DRB1], DRB3, DRB4, and DRB5) were calculated using the HLA typing result per the software at two-field resolution as a replica and integrating the seven replicates, grouped by disease states (control and JIA). The same procedure was also performed using the GSE112057 [41] grouped by disease subtypes (healthy control, oligo-JIA, poly-JIA, and sJIA) and ancestry (healthy control, African JIA, and European JIA).

The expression levels of HLA genes were estimated by seq2HLA [60] using the GSE112057 [41] dataset. FASTQ-formatted files in GSE112057 [41] were used as input files and the locus-specific expression levels of HLA genes were measured as read per kilobase million. HLA class I and II gene expression levels were compared between JIA and control groups.

Kim *et al. Arthritis Research & Therapy*   (2024) 26:47

Page 5 of 19

Profiling adaptive immune repertoires in JIA

The unmapped reads were extracted from the GSE112057 [41] dataset following a read origin protocol (ROP) [70]. SRR6868722 and SRR6868696 were excluded due to quality problems. To analyze the TCR repertoires, the reads mapped onto the complementary determining region 3 (CDR3) in TCR loci were identified using immune profiling by ROP (ImReP) [71]. ImReP assembles the clonotypes, defined as clones having identical CDR3 amino-acid sequences, and identifies the corresponding V(D)J recombination. The alpha diversity (Shannon entropy) of TCRs was measured within the immune repertoire of an individual. The alpha diversities of TCRs (TCR α and β) were respectively compared between control and JIA groups. The same analyses were conducted using the GSE112957 [41] dataset grouped by JIA subtypes (healthy controls, oligo-JIA, poly-JIA, and sJIA).

Single-cell analysis

The single-cell RNA-seq (scRNA-seq) of JIA were downloaded from NCBI SRA (SRA study: SRP288574 [72]). The 10x Genomics BAM file contains single-cell transcriptome and TCR repertoire data from single $CD4^+CD45RO^+CD25^-$ ($CD4^+$) and $CD8^+CD45RO^+$ ($CD8^+$) T cells of synovial fluid (SF) and peripheral blood (PB) tissue in seven oligo-JIA patients. The BAM files mapped to human hg19 references were converted into FASTQ files using the bamtofastq (v1.3.1) tool and remapped to human GRCh38 references (ref-2020-A). The FASTQ data were processed using the cellranger [73] (v6.1.2) count tools and analyzed using the Seurat [74] R package. To preprocess the data, we filtered out cells with either >4000 or <200 distinct features and those with >10% mitochondrial count. We conducted normalization, identification of highly variable features (i.e., feature selection), and scaling data with default settings. After performing dimensional reduction using principal component analysis (PCA), we used the Harmony [75] R package to integrate datasets derived from seven patients and two tissue types. Based on the 40 components of Harmony, we carried out Uniform Manifold Approximation and Projection (UMAP) [76] and nearest-neighbor graph construction. Then, we determined single-cell clusters using 0.5 resolution. We retrieved the results of the single-cell TCR repertoire profiling from NCBI GEO (GEO study: GSE160097).

Statistical analysis

The Bonferroni correction was applied to determine the significance thresholds of TWAS associations ($P_{TWAS} < 7.55 \times 10^{-07}$, 0.05/66,196) and PWAS associations ($P_{PWAS} < 2.16 \times 10^{-05}$, 0.05/2311). The significance thresholds

with the Bonferroni correction were also used in the LDSC-SEG analysis, heritability enrichment analysis, and genetic architecture analysis. A false discovery rate (FDR) was used to determine the significance threshold for the TWAS-GSEA, the MAGMA gene set analysis, the genetic correlation analyses, and the functional annotation analysis using DAVID. The consensus frequencies of HLA allele types were compared using two-sided 2- (control and JIA), 3- (control, African JIA, and European JIA), and 4-sample (control, oligo-JIA, poly-JIA, and sJIA) proportion tests, respectively. Using a two-tailed *t*-test, HLA class I and II gene expression levels were compared between control and JIA groups. The alpha diversities of TCRs were compared between control and JIA groups using a two-tailed *t*-test. When grouped by JIA subtypes (control, oligo-JIA, poly-JIA, and sJIA), the alpha diversities of TCRs were compared using one-way ANOVA with post hoc Tukey HSD.

Results

Identification of susceptibility genes associated with JIA risk using TWAS and PWAS

To prioritize susceptibility genes and proteins for JIA risk, we performed a TWAS by integrating JIA GWAS data with the predicted expression of 66,196 gene/tissue pairs from 12 eQTL datasets. We focused on eQTL derived from connective tissues due to JIA reflecting a chronic inflammation of connective tissues (Table S2) [4, 77]. We identified 35 significant TWAS associations across 19 genes in four independent genomic regions (1p13.2, 1q21.3, 5q11.2, and 16p11.2-12.1) ($P_{TWAS} < 7.55 \times 10^{-07}$; Fig. 1 and Table S3). Excluding three non-coding genes and one pseudogene (*NPIPB7*) from the 19 TWAS genes, 11 out of 15 (73%) genes have been suggested as JIA-associated genes, which confirms that our results were consistent with previous studies [3, 12, 78–80]. Among the remaining four genes, *MAGI3* and *NFATC2IP* were mentioned by a previous JIA TWAS study [81], while *DCLRE1B* and *NPIPB9* have not previously been emphasized as susceptibility genes for JIA, to the best of our knowledge (Fig. 1). Next, we conducted a PWAS using predictive models of plasma proteins from INTERVAL [22] and ARIC [23]. We observed two significant PWAS associations of IL27 and ERAP2 in two genomic regions (16p11.2-12.1 and 5q15) ($P_{PWAS} < 2.16 \times 10^{-05}$; Fig. 1 and Table S4), which were reportedly involved in JIA risks [12, 81].

To distinguish between genes likely causal for JIA risk from those that tag risk due to LD and shared regulatory features, we performed a probabilistic fine-mapping analysis by FOCUS [25] using the same expression weights used in our FUSION analyses. Among the 19 TWAS genes identified by FUSION, FOCUS identified
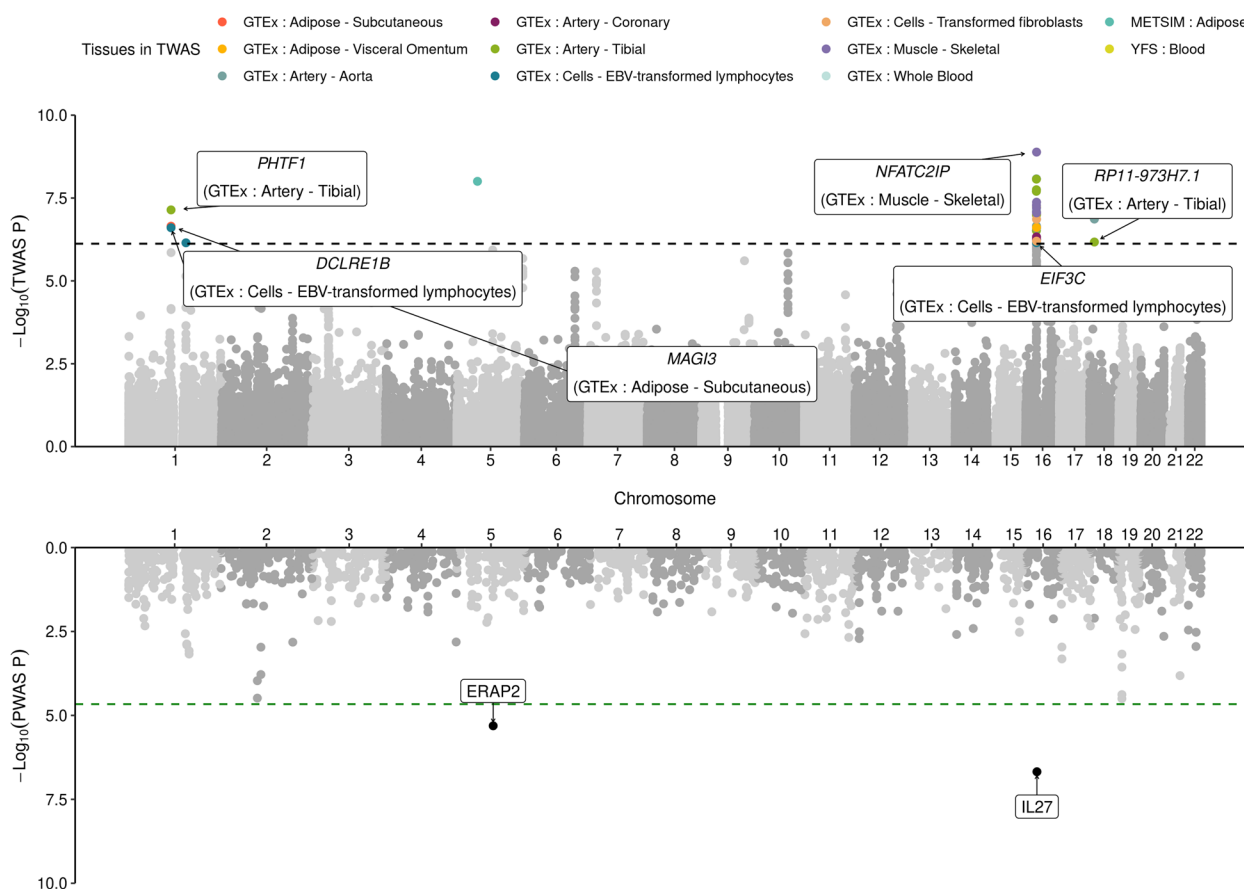
Fig. 1 Manhattan plots for JIA TWAS/PWAS result. The upper and lower panels show Manhattan plots for JIA TWAS and PWAS results, respectively. Each dot represents a *P*-value for a TWAS or PWAS association between JIA and the predicted expression level of a gene or *cis*-regulated plasma protein. The black and green horizontal dashed lines indicate the significance thresholds of TWAS and PWAS with Bonferroni correction ($P_{TWAS} < 7.55 \times 10^{-07}$ and $P_{PWAS} < 2.16 \times 10^{-05}$). Statistically significant TWAS associations of JIA are colored by tissue panels. The black dots of the lower panel indicate statistically significant PWAS associations of JIA. The names of TWAS genes and PWAS proteins with PIP > 0.2 in FOCUS are labeled

nine genes in 90%-credible sets, which we denote as putatively causal genes affected by genome-wide significant GWAS signals (Table S5). Six of the nine (67%) genes had >0.4 PIPs, of which *MAGI3* (PIP = 0.5), *NFATC2IP* (PIP = 0.417), and *DCLRE1B* (PIP = 0.405) were putatively responsible for JIA risk. Similarly, fine-mapping of PWAS results prioritized IL27 with a PIP of 0.98 in the plasma protein from the INTERVAL study (Table S5). ERAP2 from the ARIC study was included in credible sets with a PIP of 0.23. Altogether, we identified 19 susceptibility genes/ two risk proteins for JIA using TWAS/PWAS and prioritized likely causal genes for JIA risk, suggesting that our findings provide novel insights into the etiology of JIA.

Exploring biological pathways that may contribute to the pathogenesis of JIA

To explore the biological effects derived from overall TWAS associations from the multi-tissue panels, we conducted GSEA with the JIA TWAS results using CP gene sets [26]. We found a total of nine CP gene sets were significantly involved with JIA TWAS associations (FDR < 0.05). Three of the nine gene sets are characterized by sulfation (Table 1). The impaired sulfation pathway was reportedly implicated in diastrophic dysplasia leading to cartilage disorder and joint degradations, similar to the clinical manifestations of JIA [82]. In addition, tyrosine-sulfated proteins were reported to play roles in the pathogenesis of various AIDs [83]. Four other gene sets involved in the IL27 pathway, IL6 family signaling, nitric oxide 2 IL12 (NO2IL12) pathway, and IL17 pathway are associated with immune responses that are representatives of AIDs. The T cell apoptosis pathway was also significantly implicated in JIA associated with the dysregulated T cell responses [84]. Moreover, stathmin is known to play a critical role in regulating the cell cycle, and its phosphorylation is reportedly important in T cell activation [85, 86].

Kim *et al. Arthritis Research & Therapy*   (2024) 26:47

Page 7 of 19

Table 1 Biological pathways significantly involved in JIA based on the TWAS associations

Gene set	Panel	*P*-value	FDR
Wikipathway: Sulfation biotransformation reaction	GTEx: Cells - Transformed fibroblasts	4.56×10^{-07}	8.41×10^{-04}
	GTEx: Muscle - Skeletal	7.58×10^{-06}	1.12×10^{-02}
Reactome: Cytosolic sulfonation of small molecules	GTEx: Cells - Transformed fibroblasts	6.11×10^{-06}	5.63×10^{-03}
KEGG: Sulfur metabolism	GTEx: Muscle - Skeletal	1.16×10^{-05}	1.12×10^{-02}
PID: IL27 pathway	GTEx: Whole - Blood	4.97×10^{-07}	8.38×10^{-04}
Reactome: Interleukin 6 family signaling	GTEx: Whole - Blood	1.65×10^{-06}	1.39×10^{-03}
BioCarta: NO2IL12 pathway	GTEx: Whole - Blood	1.10×10^{-05}	6.17×10^{-03}
	YFS: Blood	3.09×10^{-05}	2.33×10^{-02}
BioCarta: T cell apoptosis (TCAPOPTOSIS) pathway	YFS: Blood	9.47×10^{-06}	1.85×10^{-02}
BioCarta: Stathmin pathway	YFS: Blood	3.57×10^{-05}	2.33×10^{-02}
BioCarta: IL17 pathway	YFS: Blood	5.82×10^{-05}	2.84×10^{-02}

We additionally performed a GWAS-based pathway enrichment analysis using MAGMA [33] with JIA GWAS summary statistics and CP gene sets, given that HLA signals were dropped from TWAS-based results due to complicated LD. The results showed that a total of 40 gene sets were significantly implicated in JIA (FDR < 0.05; Table S6). Most pathways were associated with various immune system components, such as inflammatory cytokines, HLA, immunoglobulins, and T cells. The IL27 pathway, IL6 family signaling, NO2IL12 pathway, and IL17 pathway were observed in both TWAS-GSEA and MAGMA (Tables 1 and S6). The gene sets representing T1D and autoimmune thyroid disease (AITD) were also related to JIA, supporting the idea that JIA is indeed an AID. Collectively, our results suggest that impaired sulfation pathways and immune signaling, especially T cell-mediated responses, may contribute to the pathogenesis mechanism of JIA.

Estimation of genetic correlations between JIA and other traits

Considering that the pathway analysis showed certain AIDs, such as T1D and AITD, were associated with JIA, we carried out genetic correlation analyses to further investigate the association of JIA with other AIDs. We estimated the genetic correlations between JIA and other traits categorized into two groups, AID-like and non-AID-like traits, at the genome- and transcriptome-wide levels. At the genome-wide level, JIA showed significant positive correlations ($FDR_{LDSC} < 0.05$) with eight out of the ten (80%) AIDs including systemic lupus erythematosus (SLE) ($r = 0.66$ and $FDR_{LDSC} = 1.24 \times 10^{-08}$) and exhibited the most significant correlation with the AID_{ALL} trait ($r = 0.53$ and $FDR_{LDSC} = 7.19 \times 10^{-11}$) (Fig. S3A). At the transcriptome-wide level, we found significant genetic overlaps ($FDR_{RHOGE} < 0.05$) between JIA

and seven AID traits including ulcerative colitis (UC), AID_{ALL} trait, Crohn's disease, rheumatoid arthritis (RA), SLE, primary biliary cirrhosis, and AID_{SURE} trait. JIA was most significantly correlated with UC ($r = 0.35$ and $FDR_{RHOGE} = 1.28 \times 10^{-06}$) and had the most positive correlation with T1D ($r = 0.65$ and $FDR_{RHOGE} = 6.91 \times 10^{-02}$) (Fig. S3B). In particular, T1D was simultaneously identified to have highly positive correlations with JIA at the genome- ($r = 0.62$ and $FDR_{LDSC} = 3.58 \times 10^{-05}$) and transcriptome-wide levels ($r = 0.65$ and $FDR_{RHOGE} = 6.91 \times 10^{-02}$). Overall, these results support the existence of shared genetic contributions to JIA and AIDs at the genome- and transcriptome-wide levels.

TWAS associations were validated by transcriptomic datasets on JIA

To validate TWAS associations are actually related to the gene expression patterns of transcriptomic data, we conducted WGCNA using a microarray dataset from JIA cases and controls (GSE13501) [40] to identify co-expression modules composed of genes with highly correlated expression patterns (see "Methods"). A total of 11 co-expression modules were detected and are listed in Table S7 (Fig. S1). First, we validated that the clustering of co-expression modules from the microarray dataset was recapitulated in independent RNA-seq datasets (GSE112057 [41] and GSE79970 [42]), respectively (*Z*-summary score > 2; Fig. S1). Next, we conducted GSEA using the 11 co-expression modules as reference gene sets and TWAS associations of each panel as a ranked gene list. Overall, we observed that TWAS associations from the GTEx visceral omentum adipose panel (normalized enrichment score (NES) = −1.93 and FDR = 2.50×10^{-02}) and aorta artery panel (NES = −1.93 and FDR = 2.40×10^{-02}) were significantly enriched in the magenta module (FDR < 0.05; Fig. S2 and Table S8). The

Kim *et al. Arthritis Research & Therapy*   (2024) 26:47

Page 8 of 19

TWAS associations from the NTR blood panel and the GTEx skeletal muscle panel were significantly involved in black (NES = 1.84 and FDR = 2.40×10^{-02}) and yellow (NES = 1.43 and FDR = 4.90×10^{-02}) modules, respectively (FDR < 0.05). These three modules were highly preserved in both RNA-seq datasets with the Z-summary score > 10 (Fig. S1). We performed functional annotation of the 11 identified modules using gene ontology (GO) terms to identify their biological functions (FDR < 0.05; Table S9) [46, 47]. The significant GO term of the magenta module was "Humoral immune response (GO:0006959)" (FDR = 4.16×10^{-02}). The black module was significantly associated with "Leukocyte migration (GO:0050900)" (FDR = 7.73×10^{-5}), while the yellow module was enriched in "Regulation of transcription, DNA-templated (GO:0006355)" (FDR = 4.34×10^{-8}) and "Transcription, DNA-templated (GO:0006351)" (FDR = 7.11×10^{-6}). Additionally, significant GO terms of the tan module were "Translational initiation (GO:0006413)" (FDR = 9.78×10^{-81}) and "rRNA processing (GO:0006364)" (FDR = 8.98×10^{-71}). Taken together, we observed that TWAS associations were actually enriched in the expression pattern of particular immunological and metabolic gene sets derived from JIA transcriptomic data, suggesting that the TWAS signals may be along with transcriptomic signals for JIA.

Disease heritability in JIA-relevant tissues and cell types

To better understand how genetic variants affect JIA risks, we aimed to identify the cell types or tissues relevant for the pathogenesis of JIA using LD score regression in specifically expressed genes (LDSC-SEG) [48]. LDSC-SEG tests whether disease heritability is enriched in regions of genes in a specific tissue using stratified LD score regression [87], analyzing transcriptome or epigenome data together with GWAS. We first applied this analysis to 53 and 152 tissues or cell types from the GTEx project [49] and Franke lab data [50, 51], respectively. Consistent with the reason for using connective tissues in the JIA TWAS, this analysis showed that lymphocytes or blood tissues were enriched in JIA (Fig. S4). Additionally, we detected enrichment for the $CD4^+$ T cells among 292 immune cell types from ImmGen [52] in JIA (Fig. S5). To support the results from the expression-based LDSC-SEG analysis, we examined whether JIA-related heritability is enriched in epigenetic markers from the ENCODE projects [54] and Roadmap Epigenomics [53]. Based on 431 tissue-specific ChIP-seq annotations of six epigenetic marks, we detected an enrichment at the 5% Bonferroni threshold ($P < 1.16 \times 10^{-04}$) for only blood tissue (Fig. S6). Notably, the most significant enrichment was observed in T cells across active promoters and gene markers (Fig. 2A) as well as active enhancer markers (Fig. 2B). To

validate the Chip-seq results, we used ATAC-seq data which is associated with chromatin accessibility in 13 blood cell types[55]. In line with the Chip-seq results, we found enrichment in T, B, and NK cells after the Bonferroni correction ($P < 3.8 \times 10^{-03}$; Fig. 2C).

To test whether these enriched tissues and cell types causally mediate JIA risk, we next estimated the proportion of heritability mediated by gene expression levels (h^2_{med}/h^2_g) in a tissue context using mediated expression score regression (MESC) [57]. When using the all-tissue meta-analyzed expression scores from GTEx, we observed 0.34 of h^2_{med}/h^2_g for JIA (standard error (se) = 0.16), which was higher than those from other AIDs ($P < 3.37 \times 10^{-02}$; Fig. 2D). We estimated lower h^2_{med}/h^2_g from the tissue-group meta-analyzed (i.e., 12 connective tissues) (h^2_{med}/h^2_g = 0.217 and se = 0.119) and individual-tissue (i.e., whole blood) expression scores (h^2_{med}/h^2_g= 0.137 and se = 0.085) than from all-tissue expression scores in GTEx data (Fig. 2D), consistent with those in previous studies using 42 diseases and complex traits[57]. We identified the highest h^2_{med}/h^2_g value in the EBV-transformed lymphocytes among 48 GTEx tissues (h^2_{med}/h^2_g = 0.189 and se = 0.067; Fig. S7), and validated the results by whole blood data from eQTLGen (Fig. 2E) [58]. Together, our findings strongly suggest that the SNP-based heritability of JIA was closely associated with gene expression and active epigenetic markers in blood tissue, especially in T lymphocytes.

Polygenicity in the genetic architecture of JIA

In complex and polygenic traits, dissecting joint distribution of effect size and MAF is important to understand the genetic architecture and to detect signals of natural selection [88]. Deleterious mutations to fitness are selected against and maintained at a low frequency by negative (purifying) selection [89, 90]. Moreover, the negative selection in autoimmune disease is important because it provides insight into the evolution of the human immune system. We conducted genetic architecture analysis using SBayesS, a recently developed method based on the Bayesian mixed linear model with GWAS summary statistics [88] to estimate the relationship between SNP effect size and MAF (S). The SBayesS method also allows us to infer multiple genetic architecture parameters including the SNP-based heritability ($h^2_{SBayesS}$) and polygenicity (π) which is the proportion of SNPs with nonzero effects. A negative value of S indicates that SNPs with lower MAF are prone to having larger effects, consistent with a model of negative selection. Overall, we estimated $S = -0.96$ (posterior standard error (p.s.e) = 0.10), providing evidence that the genetic variants related to JIA have been under negative selection (Fig. 3A). Excluding SNPs in the HLA region
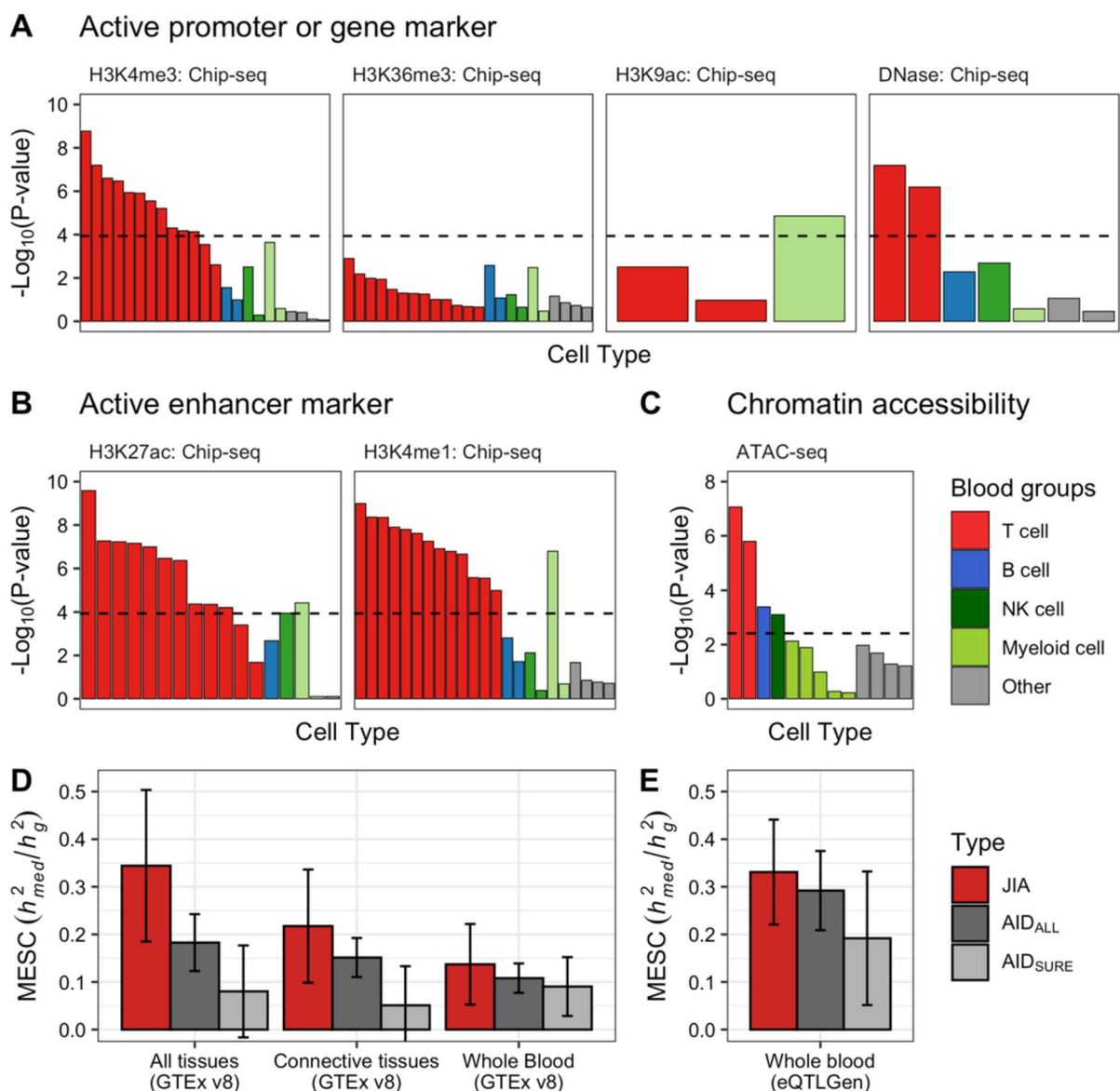
Kim *et al. Arthritis Research & Therapy* (2024) 26:47

Page 9 of 19

Fig. 2 Disease heritability analysis of JIA. **A–C** LD score regression in specifically expressed genes (LDSC-SEG) analysis applied to JIA GWAS data using epigenetic markers from blood cell types. **A** The enrichment results of LDSC-SEG analysis using active promoter or gene markers. **B** The enrichment results of LDSC-SEG analysis using active enhancer markers. The black dotted line represents a significant threshold based on the Bonferroni-corrected $P < 1.16 \times 10^{-04}$ (0.05/431). **C** The enrichment results of LDSC-SEG analysis using ATAC-seq data. The black dotted line represents a significant threshold based on the Bonferroni-corrected $P < 3.85 \times 10^{-03}$ (0.05/13). **D,E** Estimation of the proportion of heritability mediated by the gene expression levels (h^2_{med}/h^2_g) using mediated expression score regression (MESC) for JIA, AID_{ALL}, and AID_{SURE}. **D** The MESC results of all tissues (expression scores from meta-analyses across all 48 GTEx tissues), connective tissues (expression scores from meta-analyses using 12 connective tissues), and whole blood tissue from GTEx v8. **E** The MESC results from whole blood tissue from eQTLGen. Error bars indicate jackknife standard errors

(chr6:28-34Mb) increased the estimate S in JIA ($\widehat{S} = -0.49$ and p.s.e $= 0.2$), which suggests that the SNPs in the HLA region contributing toward JIA risk may have been under negative selection. The previous studies showed that the majority of AIDs have genetic relationships with the HLA area, and numerous distinct HLA alleles can predispose people to AIDs [91, 92]. For the SNP-based heritability, the estimate of $h^2_{SBayesS}$ for JIA was 0.47 (p.s.e $= 0.03$), which was higher when compared with other AIDs (Fig. 3A). We also observed that excluding the HLA region reduced the $h^2_{SBayesS}$ to 0.34 in JIA, which is consistent with previous studies in AIDs [3, 93].
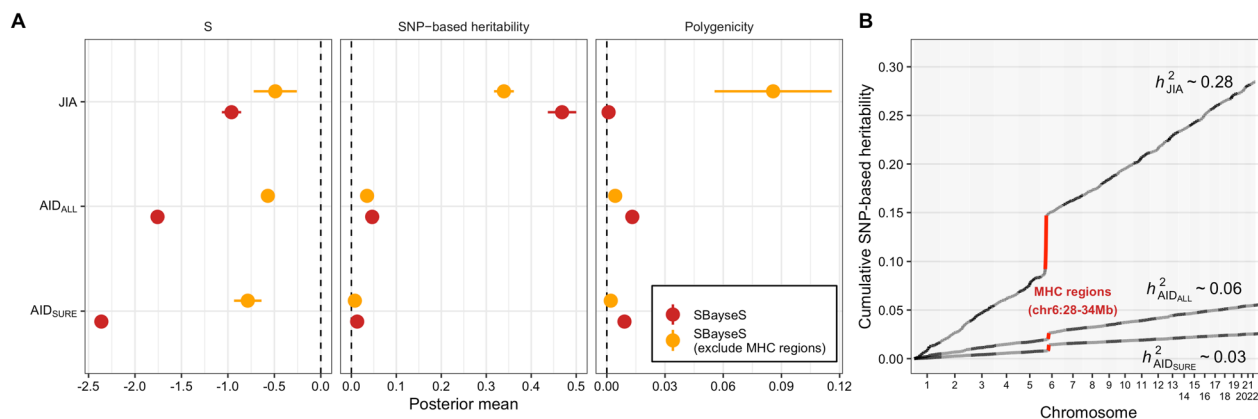
Fig. 3 Genetic architecture of JIA. **A** Estimation of the three genetic architecture parameters for JIA, AID$_{ALL}$, and AID$_{SURE}$ using Summary-data-based BayesS (SBayesS). The dots and horizontal bars represent the posterior means and standard errors, respectively. The red and orange colors represent SBayesS and SBayesS excluding HLA regions. **B** Cumulative local SNP heritability across the genome using heritability estimates from Summary statistics (HESS). Total SNP-based genes are indicated. The red color shows SNP heritability explained by the HLA region

Importantly, the polygenicity increased to 8.57% after excluding the HLA region, although the estimated polygenicity (π) is about 0.09% in JIA.

Moreover, we estimated the SNP heritability (h^2_{HESS}) using HESS [39] in each of 1703 independently partitioned genomic LD blocks across the genome. The results revealed that the total h^2_{HESS} was 0.28 in JIA (Fig. 3B), slightly lower than that based on SBayesS ($h^2_{SBayesS}$) (Fig. 3A), and the HLA region significantly explained a total of 17.5% heritability in JIA (Bonferroni-corrected $P < 2.94 \times 10^{-05}$; Fig. 3B). In line with the $h^2_{SBayesS}$ results, the total h^2_{HESS} in JIA was higher than the other AIDs. Aside from the HLA region, most of the SNP-based heritability was distributed uniformly across the genome in JIA and other AIDs. Collectively, these results strongly suggest that JIA shows higher polygenicity than other AIDs in both HLA regions and outside of HLA regions, and SNP-based heritability comes from a vastly polygenic background.

Estimation of the consensus frequencies of HLA allele types and HLA gene expression

The HLA genes in the HLA region on chromosome 6p21 encode several essential proteins in the immune system [94]. Consistent with the SNP-based heritability results in the JIA (Fig. 3B), variants in the HLA regions explain more heritability than many other variants for many diseases [95–97]. Identifying HLA allele type is essential to better understanding the disease etiology, and many tools have been developed for HLA typing using NGS data [42, 60–66]. To explore the association between HLA type diversity and JIA risk, we estimated consensus HLA allele type frequencies at 12 major loci using seven HLA-typing software (see "Methods"). We focused on the allele

types with a significantly different distribution between healthy subjects ($n = 12$) and JIA patients ($n = 115$) ($P < 0.05$; Table S10). At HLA class I loci, the frequency of A*03:01 was approximately tripled in JIA patients (12.89%) compared with that in healthy subjects (4.22%) ($P = 1.10 \times 10^{-03}$; Fig. S8). At HLA class II loci, the frequencies of DPB1*04:01, DQB1*03:02, and DRB1*01:01 were increased by more than twofold in the JIA group (33.46%, 14.39%, and 10.29%) compared to in the control group (11.21%, 4.23%, and 4.17%), respectively ($P_{DPB1*04:01} = 8.99 \times 10^{-07}$, $P_{DQB1*03:02} = 7.24 \times 10^{-04}$, and $P_{DRB1*01:01} = 1.83 \times 10^{-02}$). Notably, DRB1*04:01 (11.39%) and DRB3*02:01 (11.43%) were only detected in JIA patients. When grouped into three JIA subtypes (oligo-JIA, poly-JIA, and sJIA) (Fig. S9 and Table S11), the frequencies of DRB1*04:01 and DRB3*02:01 were the highest in poly-JIA patients. Additionally, for the transcriptomic analysis of HLA regions, we estimated the locus-specific expression levels of nine HLA genes in JIA patients and healthy controls. The results showed that the expression levels of *HLA-DPA1*, *DPB1*, *DQB1*, and *DRA* were significantly lower in JIA patients than in healthy controls ($P < 0.05$; Fig. S10). HLA class II is known to be strongly associated with susceptibility to many AIDs including JIA [98–100]. Although various mechanisms could induce the downregulation of HLA class II gene expression, the lowered expression could result in diminished tolerance induction of self-reactive T cells, leading to AIDs in the end [101].

In the HLA region, haplotypes are specific to an individual ancestral population due to the population-specific positive selection [102]. We calculated the consensus frequencies of HLA types of JIA in African and European ancestry to investigate whether heterogeneity is derived from different ancestry in HLA allele types for JIA. We

Kim *et al. Arthritis Research & Therapy*   (2024) 26:47

Page 11 of 19

focused on the allele types with significantly different distributions between healthy controls ($n = 12$), African JIA ($n = 34$), and European JIA patients ($n = 81$) ($P < 0.05$; Table S12). At HLA class I loci, B*45:01 was not detected in healthy controls and was more frequent in African JIA patients (11.18%) than European JIA patients (0.62%) ($P = 8.33 \times 10^{-22}$; Fig. 4A). B*40:01, which was not observed in African patients, had higher frequencies in European patients (10.65%) than in healthy controls (5.36%) ($P = 4.90 \times 10^{-02}$). At HLA class II loci, the frequencies of DQA1*01:01 and DQA1*03:01 were increased by more than fourfold in the European JIA patients (13.68% and 12.84%) compared to in the African JIA patients (3.05% and 1.78%), respectively ($P_{DQA1*01:01} = 9.65 \times 10^{-08}$, $P_{DQA1*03:01} = 2.56 \times 10^{-09}$; Fig. 4B). The frequencies of DRB1*04:01 observed only in JIA patients were higher in European patients (15.10%) than African patients (2.49%) ($P = 1.02 \times 10^{-09}$; Fig. 4B). Collectively, our results showed that HLA allele types had significantly imbalanced distributions between the JIA and control groups as well as between African and European ancestry, which suggests that they may be involved in JIA etiologies.

The T cell receptor repertoire reveals clonal relationships between different subpopulations

Along with HLA molecules, antigen-experienced memory T cells have been implicated as critical drivers of autoimmune inflammation [103–106]. Consistent with these previous studies, we suggested that T cells were associated with JIA etiology using TWAS and JIA SNP-heritability analysis. Therefore, we hypothesize that the TCR repertoire may be involved in JIA etiology because the TCRs mediate the recognition of HLA and provide critical insights into the adaptive immune response in health and disease [107]. To dissect whether the TCR repertoire is related to JIA, we estimated the locus-specific alpha diversities of TCR CDR3 reads using the total number of distinct clonotypes and their relative frequencies at the bulk RNA-seq level (GSE112057 [41]). We observed alpha diversities of TCR α and β were significantly decreased in JIA patients compared with controls ($P_{TCR\alpha} = 9.60 \times 10^{-04}$ and $P_{TCR\beta} = 1.36 \times 10^{-02}$), which indicates that the clonotypic diversities of TCRs were reduced in the JIA group (Fig. 5A). When JIA was grouped into three subtypes (oligo-JIA, poly-JIA, and sJIA), the patterns of alpha diversity in all subtypes were consistent with those in the JIA group (Fig. S11 and Table S13). These results suggest that a few clonotypes positively affecting the development of JIA may be dominant in the immune repertoire of JIA patients.

Next, we investigated whether different T cell populations showed differences in TCR diversity at the single-cell level. Single-cell TCR sequencing can accurately measure the diversity of T cell populations, which is critical for understanding the complexity of the immune response by providing paired TCR α and β information [108]. We used scRNA-seq data (SRP288574 [72]) derived from single CD4+CD45RO+CD25− (CD4+) and CD8+CD45RO+ (CD8+) T cells in SF and PB tissues of seven oligo-JIA patients. After a series of quality control filters (see "Methods"), nine CD4+ and seven CD8+ T cell clusters were identified (Fig. S12, S13, and 5B), and each cluster represented a distinct distribution of clonotypes (Fig. 5C). Among 67,235 single T cells, 33,855 cells (50.3%) had at least one pair of full-length TCR α and β chains. In addition, 13,113 of the 33,855 cells (38.7%) expressed a full-length α-β chain pair detected at least twice. Expanded clonotype cells were most prevalent in two CD8+ clusters (CD8_C1 and CD8_C3) and one CD4+ cluster (CD4_C5) (Fig. 5C). Compared with the other clusters, the CD8_C1 cluster had higher expression of several markers of recently activated effector memory or effector T cells (designated as CD8+ *GZMH*+ T_{EMRA} cells) and the CD8_C3 cluster had higher expression of several markers of effector memory T cells (designated as CD8+ *GZMK*+ T_{EM} cells) (Table S14). The CD4_C5 cluster showed higher expression of *CXCL13* and *BHLHE40* (designated as *CXCL13*+*BHLHE40*+ T_H cells) (Fig. S14 and Table S14). Each cluster of the CD8+ *GZMH*+ T_{EMRA}, CD8+ *GZMK*+ T_{EM}, and *CXCL13*+*BHLHE40*+ T_H cells was observed to have a substantially lower alpha diversity value than most other clusters (Fig. 5D). Additionally, the single T cell analysis by RNA-seq and TCR tracking expansion (STARTRAC-expa) index, which quantitatively describes tissue clonal expansion, revealed the CD8+ *GZMH*+ T_{EMRA}, CD8+ *GZMK*+ T_{EM}, and *CXCL13*+*BHLHE40*+ T_H cells as the clusters with the highest degree of clonal expansion for each cell type (Fig. S15).

To emphasize disease-specific memory T cell clusters, we conducted GSEA using the co-expression modules from case-control JIA expression data as reference gene sets (Fig. S16). The results showed that only the *CXCL13*+*BHLHE40*+ T_H cells were positively enriched in the black module gene set (NES = 1.93 and FDR = 1.20×10^{-6}; Fig. 5E). Notably, the black module genes, associated with leukocyte migration essential for inflammation and innate immunity [109] (Table S9), were positively enriched in TWAS associations from NTR blood tissue (Fig. S2C). In summary, we identified clonally expanded T cell subpopulations in JIA patients. The *CXCL13*+*BHLHE40*+ T_H cells were significantly related to case-control JIA expression data, suggesting that the cells might be potential therapeutic targets for JIA.
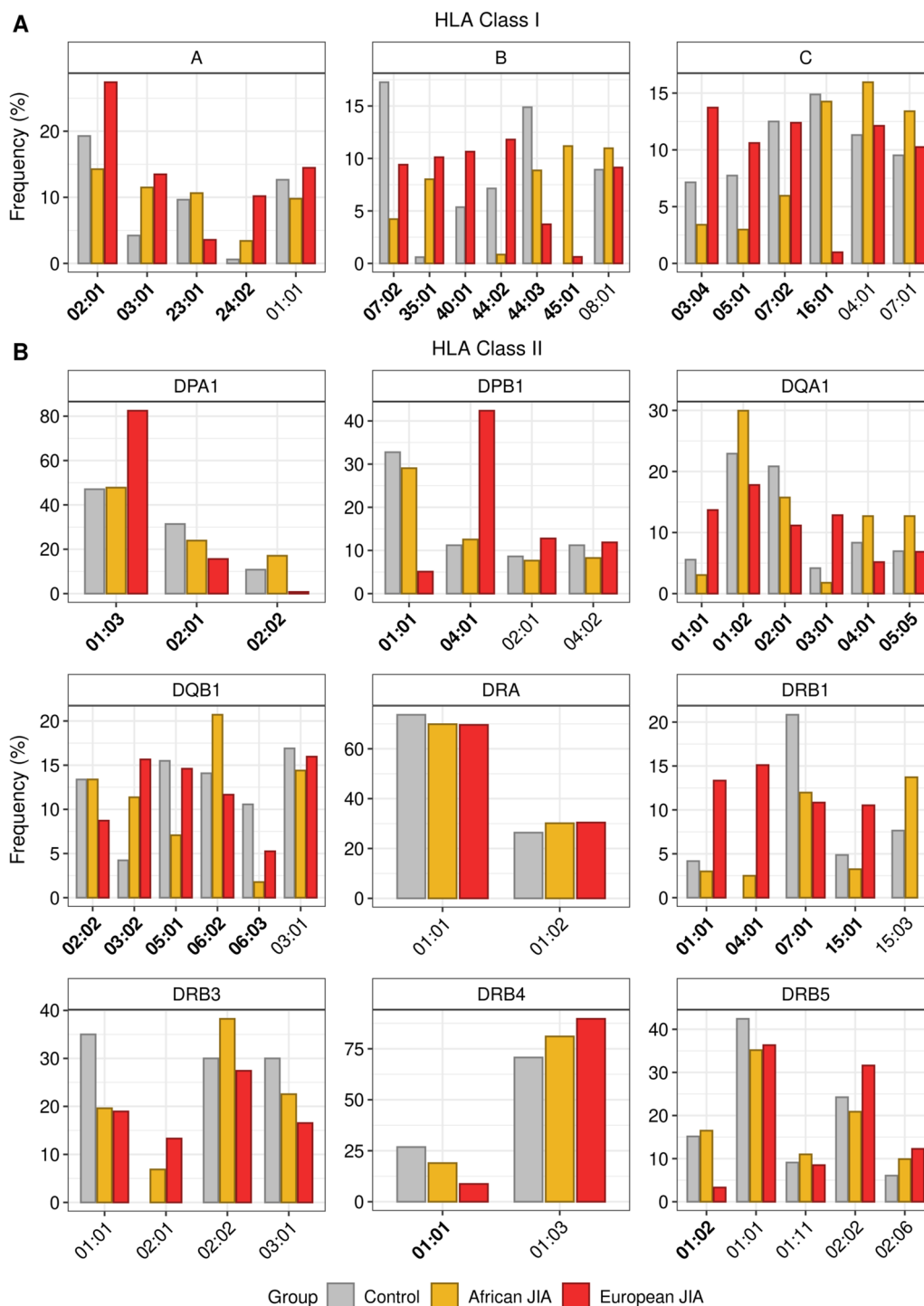
Fig. 4 Consensus frequencies of HLA allele types in healthy controls, African JIA patients, and European JIA patients. **A** Bar plots showing the consensus frequencies of HLA allele types at HLA class I loci. **B** Bar plots showing the consensus frequencies of HLA allele types at HLA class II loci. The *x*- and *y*-axis indicate the names and frequencies of HLA allele types, respectively. The names of HLA types having significantly different distributions between healthy controls, African JIA patients, and European JIA patients were represented in bold (*P* < 0.05)

Kim *et al. Arthritis Research & Therapy* (2024) 26:47

Page 13 of 19

Discussion

Using a TWAS/PWAS to identify biologically interpretable susceptibility genes/proteins, we detected 19 TWAS genes and two PWAS proteins significantly associated with JIA risks (Fig. 1). Since trait-associated genes/proteins identified by TWAS/PWAS do not fully elucidate the disease's causality [110], we performed a fine-mapping analysis to prioritize putatively causal genes/proteins for JIA. We found that *MAGI3*, *DCLRE1B*, *NFATC2IP*, IL27, and ERAP2 were responsible for JIA risks and are implicated in the immune system. *MAGI3* encodes the PDZ proteins involved in T cell homeostasis and mediates the suppression of the PI3K/Akt pathway [111, 112]. The downregulation of *MAGI3* (Z_{TWAS} = −5.18; Table S3) may lead to an upregulation of the PI3K/Akt pathway, which, in turn, downregulates the differentiation of T cells toward Treg [10, 112]. *DCLRE1B* was reportedly implicated in an inherited bone marrow failure syndrome associated with immune deficiency [113]. *NFATC2IP* regulates the nuclear factor of activated T cells (NFAT)-driven transcription of specific cytokine genes, including *IL4* in T-helper 2 cells [114]. Considering that IL4 production is affected by IL27 [115], *NFATC2IP* may be linked to the inflammatory cytokine pathways identified by the TWAS-GSEA (Table 1). Consistent with our PWAS result of IL27 (Z_{PWAS} = −5.19; Table S4), the synovial fluid level of IL27 was reported to be significantly decreased in enthesitis-related arthritis (ERA) patients, one of the JIA subtypes [116]. Additionally, ERAP2 (Z_{PWAS} = 4.57; Table S4) reportedly has crucial roles in immunomodulating immune responses [117]. The SNP in a splice site for *ERAP2* had a genome-wide significant association with JIA [99], and polymorphism in genes encoding ERAP1 and ERAP2 is known to predispose to ERA [118]. While our TWAS genes, *MAGI3* and *NFATC2IP*, were mentioned in previous JIA TWAS studies [81, 119, 120], we could better understand the genetic contribution of risk genes to JIA pathogenesis by conducting a PWAS together with fine-mapping and TWAS-based pathway enrichment analysis. A proteome-level analysis can provide more relevant biological information, capture alternative splicing, and post-translational modifications about disease mechanisms.

In genetic architecture analysis, SNP-based heritability was spread uniformly throughout the genome aside from a modest fraction in the HLA regions (about 18%) (Fig. 3). The results can be explained by an "omnigenic model". The omnigenic model is a theoretical framework in genetics and genomics that proposes that most traits, diseases, and other complex phenotypes are influenced by the combined effects of a large number of genes, rather than being driven by just a few "core" genes [121, 122].

Alongside the TWAS/PWAS and genetic architecture analysis, we conducted HLA-typing analyses as per the disease state and ancestry. In the investigation of JIA susceptibility, HLA typing is an indispensable tool for pinpointing genes and proteins linked to the disease, becoming essential in immunogenetics research [123]. Different human populations exhibit different genetic architecture due to diverse LD patterns, and studying multiple ancestries allows for a more comprehensive understanding of HLA variation. Additionally, the genes encoding HLA are characterized by a remarkable degree of polymorphism, meaning they exist in numerous different HLA alleles. The prevalence of these alleles varies considerably across ethnic groups [124, 125]. Understanding HLA typing among various ancestries can have important clinical consequences, opening doors to more precise medical diagnoses and targeted therapeutic interventions. As the rigorous HLA analysis using seven HLA-typing software, we suggested DRB1*04:01 and DRB3*02:01 showing >10% consensus frequencies in only JIA patients as risk alleles for JIA. In line with our result, it was reported that DRB1*01 and DRB1*04 might be implicated in the genetic predisposition of rheumatoid factor+ JIA and that DRB1*04 was confirmed to be involved in sJIA [126]. In addition, DRB1*04:01 and DQB1*03:02 may have a shared contribution to JIA and T1D since specific interactions between DRB1*03:01-DQB1*02:01/DRB1*04:01-DQB1*03:02 genotypes were previously described to increase T1D risks [127]. As

---

(See figure on next page.)

Fig. 5 TCR diversities in JIA. **A** Box plots showing the alpha diversities of TCR α and β in healthy control and JIA groups. The *y*-axis indicates alpha diversity representing the clonotypic diversity of specific TCR locus. Green and red dots indicate samples of healthy controls and JIA patients, respectively. Asterisks denote the significance levels of differences between the clonotypic diversities of TCRs within healthy controls and that within JIA patients. *, *P* < 0.05; **, *P* < 0.01; ***, *P* < 0.001. **B** A Uniform Manifold Approximation and Projection (UMAP) plot of 67,235 T cells from seven JIA patients showing 16 major clusters (nine for 34,605 CD4[+] and seven for 32,630 CD8[+] T cells). Each dot represents an individual T cell and color indicates cluster origin. **C** Left. A bar plot showing the number of single T cells and frequencies of unique and expanded T cell clones in each cluster. The inset numbers indicate the proportion of the cell type in the cluster. Right. UMAP plot showing the clusters with the clone types. The colors correspond to the number of clones (i.e. clonal abundance). **D** A scatter plot showing the alpha diversity of TCR in each cluster. The red and blue colors denote CD4[+] and CD8[+] T cells, respectively. **E** Module enrichment analysis between expression levels of scRNA-seq CD4_C5 cluster and black co-expression gene set derived from JIA case-control expression data (GSE13501). The gene list of the black modules is in Table S9.
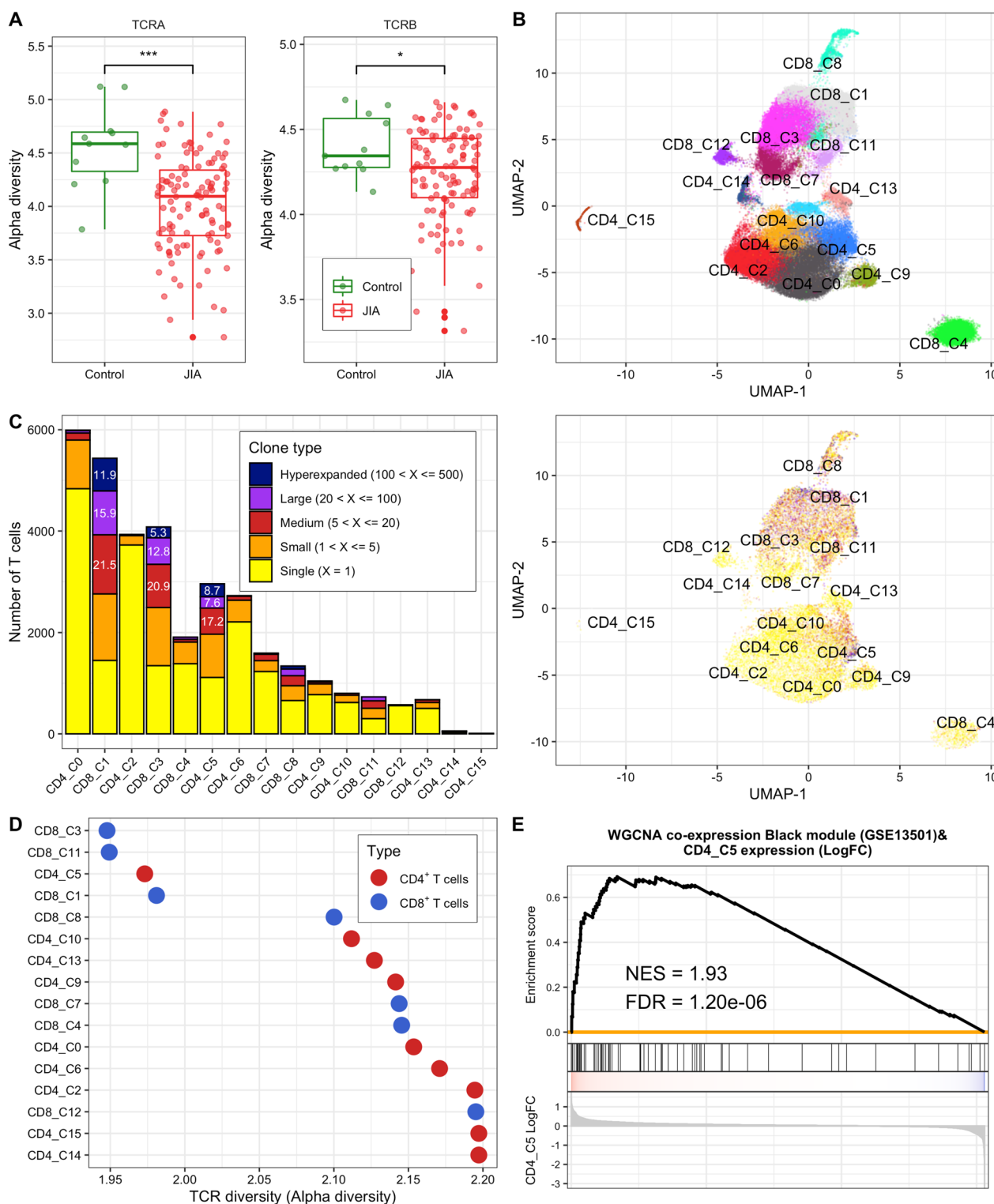
Fig. 5 (See legend on previous page.)

the distribution of HLA alleles varies among different ancestries, multi-population needs to be considered an important factor in studying the associations between

HLA allele types and disease risks [128]. Through HLA-typing analysis utilizing JIA patients' multi-ancestry information, we identified ancestry-specific risk alleles in

Kim *et al. Arthritis Research & Therapy*   (2024) 26:47

Page 15 of 19

both HLA class I (B*40:01 and B*45:01) and HLA class II (DQA1*01:01, DQA1*03:01, and DRB1*04:01) (Fig. 4). Even though some HLA types were previously reported as risk alleles for JIA in specific populations, few HLA studies have compared risk alleles for JIA between different ancestral groups [123, 129].

A recent study showed that there is a hypothesis that HLA risk alleles may affect the risks of autoimmunity by influencing thymic T cell selection [130]. T cells with receptors that recognize self-HLA molecules and interact with foreign antigens are selected during T cell development in the thymus [131]. In AIDs, self-antigens may induce immune responses by self-reactive T cells similar to how foreign antigens trigger immune responses [132]. As the proliferation of antigen-specific lymphocytes is induced by immune responses, we believe that the reduced alpha diversities of TCRs within JIA patients may be attributed to the selective increase of a few different self-reactive clonotypes at the levels of individual patients (Fig. 5). In fact, circulating $CD4^+$ T cells replicating the phenotypical signature of T lymphocytes infiltrating the inflamed synovium were increased in patients with JIA [133]. Previous clinical studies also reported that the alpha diversity of the TCR repertoire was significantly reduced in patients with RA or SLE that are also AIDs [134–136]. At the single-cell level, our results showed that clonally expanded T cell subpopulations in JIA patients, especially $CXCL13^+BHLHE40^+$ T_H cells, were significantly involved in JIA risks (Fig. 5). The $PD-1^+TOX^+BHLHE40^+$ population of $CD4^+$ T cells was reported to presumably support extrafollicular B cell activation by secreting IL21 and CXCL13 in JIA [72]. In addition, it was reported that CXCL13-producing $CD4^+$ T_H cells induced in RA synovium may be involved in the recruitment of B cells and circulating follicular helper T cells at inflammation sites [137].

Although our study successfully identified susceptibility genes/proteins for JIA by TWAS/PWAS, functional studies are needed to clarify the exact genetic effects derived from the genes/proteins. We excepted genes from the HLA region due to its structural diversity and long-range LD in the TWAS/PWAS; however, the GWAS trait associations have been more reported in the HLA region than in any other locus [11]. To compensate for this situation, we identified the consensus HLA allele types using seven different HLA-typing tools with RNA-seq data. The genetic architecture analysis, especially for SNP-based heritability, requires additional confirmation because we estimated the heritability using GWAS summary statistics data with reference LD dataset. In addition, the publicly available data we used had some limitations. While it is crucial to study the pathological differences between JIA subtypes, the GWAS summary statistics data for JIA comprising 8 subtypes was used because there were very few publicly available GWAS data for each JIA subtype (See "Methods"). The scRNA-seq dataset was derived solely from oligo-JIA patients. Since the RNA-seq dataset for HLA typing did not contain ancestral information on healthy subjects, the ancestry-specific risks of HLA alleles need to be further confirmed in the control group. Additionally, most eQTL/pQTL datasets for TWAS and PWAS, along with GWAS summary statistics for JIA, are primarily focused on individuals of European descent. This creates a critical gap in our understanding of complex traits across diverse populations [138]. Fortunately, recent advancements in comprehensive public resources have enabled researchers to employ multi-ancestry approaches for TWAS/PWAS [23, 139], providing a way for more inclusive and robust genetic analyses for JIA.

Conclusions

Our findings shed new light on the pathogenesis of JIA and provide a strong foundation for future mechanistic studies aimed at uncovering the molecular drivers of JIA.

Abbreviations

JIA	Juvenile idiopathic arthritis
AID	An autoimmune disease
SNP	Single-nucleotide polymorphism
GWAS	Genome-wide association study
eQTL	Expression quantitative trait loci
pQTL	Protein quantitative trait loci
TWAS	Transcriptome-wide association study
PWAS	Proteome-wide association study
HLA	Human leukocyte antigen
TCR	T cell receptor
LD	Linkage disequilibrium
MHC	Major histocompatibility complex
PIP	Posterior inclusion probability
T1D	Type 1 diabetes
UC	Ulcerative colitis
SLE	Systemic lupus erythematosus
RA	Rheumatoid arthritis
FDR	False discovery rate
SF	Synovial fluid
PB	Peripheral blood
ERA	Enthesitis-related arthritis

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13075-024-03280-2.

Additional file 1: Supplementary Figure S1. Identification of co-expression modules using GSE13501 and WGCNA. (A) A dendrogram showing co-expression modules based on the dissimilarity of topological overlap measurement. Color bars represent the randomly assigned colors for the module names. The orders of module colors are purple, green, black, magenta, yellow, turquoise, blue, red, pink, tan, and brown. (B) The result of module preservation analysis using GSE13501 as a reference set and GSE112057 as a test set. (C) The result of module preservation analysis using GSE13501 as a reference set and GSE79970 as a test set. The green and blue dotted lines indicate the thresholds of significance, respectively (Z-summary score > 2 and Z-summary score > 10). **Supplementary**

Figure S2. Functional annotation of co-expression modules enriched with TWAS associations. (A) A GSEA plot using a ranked gene list from GTEx: Adipose Visceral Omentum and magenta module (normalized enrichment score (NES) = −1.93 and FDR = 0.025) (B) A GSEA plot using a ranked gene list from GTEx: Artery Aorta and magenta module (NES = −1.93, and FDR = 0.024). (C) A GSEA plot using a ranked gene list from NTR: Blood and black module (NES = 1.84, and FDR = 0.024). (D) A GSEA plot using a ranked gene list from GTEx: Muscle Skeletal and yellow module (NES = 1.43, and FDR = 0.049). **Supplementary Figure S3.** Genetic correlations between JIA and other traits at the genome- and transcriptome-wide levels. AID-like and non-AID-like traits were compared with JIA. The range of genetic correlation coefficient and FDR values are represented by a color bar and symbol size, respectively. The shape of the symbol indicates the significance of the corresponding correlation. (A) Genetic correlations between JIA and other traits at the genome-wide level. (B) Genetic correlations between JIA and other traits at the transcriptome-wide level, based on the TWAS results. S stands for significant and NS for non-significant. **Supplementary Figure S4.** Heritability enrichment analysis using tissue or cell type expression. Linkage disequilibrium (LD) score regression in specifically expressed genes (LD-SEG) analysis applied to JIA GWAS data. (A) The enrichment results of LD-SEG analysis using GTEx data. (B) The enrichment results of LD-SEG analysis using Franke lab data. The red dotted lines represent a significant threshold based on the Bonferroni correction. P-values of 9.43×10^{-04} (0.05/53) and 3.28×10^{-04} (0.05/152) are the significant thresholds of the results from GTEx and Franke lab data, respectively. **Supplementary Figure S5.** Heritability enrichment analysis using immune cell type expression. The enrichment results of LD-SEG analysis using ImmGen datasets. The black dotted line represents a significant threshold based on the Bonferroni-corrected $P < 1.71 \times 10^{-04}$ (0.05/292). **Supplementary Figure S6.** Heritability enrichment analysis using epigenetic markers. The enrichment results of LD-SEG analysis using epigenetic markers of different tissue types. The red dotted lines represent a significant threshold based on the Bonferroni-corrected P-value$< 1.16 \times 10^{-04}$ (0.05/431). **Supplementary Figure S7.** Estimation of the proportion of heritability mediated by the gene expression levels (). The bar plots show the results of mediated expression score regression (MESC) using GTEx v8 and eQTLGen data. Error bars indicate jackknife standard errors. **Supplementary Figure S8.** Consensus frequencies of HLA allele types in healthy control subjects and JIA patients. (A) Bar plots showing the consensus frequencies of HLA allele types at HLA class I loci. (B) Bar plots showing the consensus frequencies of HLA allele types at HLA class II loci. The x- and y-axis indicate the names and frequencies of HLA allele types, respectively. The names of HLA types detected only in the JIA group are marked in red. The names of significantly different HLA types observed in the healthy control and JIA groups are represented in bold ($P < 0.05$). **Supplementary Figure S9.** Consensus frequencies of HLA allele types in healthy control subjects and JIA patients grouped by subtypes. (A) Bar plots showing the consensus frequencies of HLA allele types at HLA class I loci. (B) Bar plots showing the consensus frequencies of HLA allele types at HLA class II loci. The x- and y-axis indicate the names and frequencies of HLA allele types, respectively. The names of significantly different HLA types observed in the healthy control and 3 JIA subtype groups are represented in bold ($P < 0.05$). **Supplementary Figure S10.** Boxplots showing the locus-specific expression levels of HLA class II genes. Boxplots showing the expression levels of HLA class II genes, (A) *DPA1*,(B) *DPB1*, (C) *DQB1*, and (D) *DRA*, in healthy controls and JIA patients. Asterisks represent the significance levels of difference between the HLA gene expression levels in JIA patients and those in healthy controls. *, $P < 0.05$; **, $P < 0.01$. **Supplementary Figure S11.** The clonotypic diversities of TCRA and TCRB loci in healthy controls and patients of 3 JIA subtypes. Box plots showing the alpha diversities of (A) TCRA and (B) TCRB loci in the healthy controls and 3 JIA subtypes. The y-axis indicates alpha diversity representing the clonotypic diversity. Colored dots indicate samples of healthy controls and JIA patients of specific subtypes, respectively. Asterisks represent the significance levels of differences between the clonotypic diversities of TCRA and TCRB in healthy controls and those in patients of each JIA subtype. **Supplementary Figure S12.** Quantification of the cluster distribution.

(A) Bar plot showing cluster distribution of the two different T cell types. (B) Bar plot showing cluster distribution of the seven different JIA patients. (C) Bar plot showing cluster distribution of the two tissue types. The y-axis of the bar graph denotes the normalized proportions. (D) Alluvial diagram showing the distribution of T cell types. Three categorical axes are variables (i.e., origin tissues, clusters, and T cell types) along which the data are grouped. Each horizontal spline (called alluvium) corresponds to a fixed value of each axis variable, indicated by T cell types' of colors. The red and blue colors correspond to CD8$^+$ and CD4$^+$ T cells, respectively. **Supplementary Figure S13.** UMAP plot for T cell types and tissue types in different clusters. Each dot represents an individual T cell and color indicates cluster origin. **Supplementary Figure S14.** Expression levels of signature genes in each cluster. (A) UMAP plot showing the expression levels of four selected genes. Each hexagon represents summarizing points into binned hexagon cells (The number of bins partitioning the range = 50) using schex R package. The color bar denotes the proportion of observations in the bin greater than 0. (B) Dot plot representing the expression levels of six selected genes. The size of dots denotes the percentage of cells within a cluster and the color bar encodes the average expression level of the selected genes across all cells within a cluster. Supplementary Figure S15. Clonal expansion levels of the clusters. Clonal expansion levels of each cluster in (A) CD4$^+$ T cells and (B) CD8$^+$ T cells. STARTRAC-expa quantified clonal expansion levels for each JIA patient ($n = 7$). *FDR < 0.05, **FDR < 0.01, two-sided Wilcoxon test. **Supplementary Figure S16.** Module enrichment analysis between expression levels of scRNA-seq CD4_C5 cluster genes and co-expression gene sets derived from JIA case-control expression data. GSEA plots between expression levels of scRNA-seq CD4_C5 cluster and co-expression of (A) Tan and (B) Green modules derived from case-control expression data of GSE13501. The gene lists of modules are in Table S7.

Additional file 2: Table S1. The list of 12 reference eQTL panels used in the TWAS for JIA. **Table S2.** The list of the total TWAS associations for JIA from the connective tissues. **Table S3.** The list of 35 significant TWAS associations for JIA (PTWAS < 7.55×10^{-07}). **Table S4.** The list of the total PWAS associations for JIA. Asterisks represent significant PWAS associations after Bonferroni correction (PPWAS < 2.16×10^{-05}). **Table S5.** The result of fine-mapping of TWAS/PWAS signals by using the FOCUS. **Table S6.** Biological pathways significantly involved in JIA, identified by the MAGMA (FDR < 0.05). **Table S7.** The list of Entrez IDs of genes in 11 co-expression modules identified by WGCNA. **Table S8.** The result of gene set enrichment analysis with TWAS associations using co-expression modules as reference gene sets. **Table S9.** Functional annotation of the 11 modules with GO terms using DAVID, respectively. **Table S10.** HLA allele types having significantly imbalanced distribution between healthy controls and JIA patients ($P < 0.05$). **Table S11.** HLA allele types having significantly imbalanced distribution between healthy controls, oligo-JIA, poly-JIA, and sJIA patients ($P < 0.05$). **Table S12.** HLA allele types having significantly imbalanced distribution between healthy controls, African JIA, and European JIA patients ($P < 0.05$). **Table S13.** The P-values calculated by comparing the alpha diversity of TCRs. **Table S14.** Expression profiles of gene markers in each T cell cluster.

Authors' contributions
DK and JJ designed the study. NM and SM provided advice on some aspects of the study methods. DK, JS, and JJ performed the data analysis. DK, JJ, and WJ wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Availability of data and materials
Not applicable.

Kim *et al. Arthritis Research & Therapy*   (2024) 26:47

Page 17 of 19

Declarations

Author details
[1]Department of Life Sciences, Dongguk University-Seoul, Seoul 04620, Republic of Korea. [2]Center for Genetic Epidemiology, Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. [3]Department of Quantitative and Computational Biology, USC Dornsife College of Letters, Arts and Sciences, University of Southern California, Los Angeles, CA, USA. [4]Titus Family Department of Clinical Pharmacy, USC Alfred E. Mann School of Pharmacy and Pharmaceutical Sciences, University of Southern California, Los Angeles, CA, USA. [5]Department of Computational Biomedicine, Cedars-Sinai Medical Center, Hollywood, CA, USA.

References

1. Barut K, Adrovic A, Şahin S, Kasapçopur Ö. Juvenile Idiopathic Arthritis Balk. Med J. 2017;34:90–101.
2. Kim D, Song J, Lee S, Jung J, Jang W. An integrative transcriptomic analysis of systemic juvenile idiopathic arthritis for identifying potential genetic markers and drug candidates. Int J Mol Sci. 2021;22:712.
3. Li J, et al. Identification of Target Genes at Juvenile Idiopathic Arthritis GWAS Loci in Human Neutrophils. Front Genet. 2019;10:181.
4. Gusev A, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet. 2016;48:245–52.
5. Shiina T, Hosomichi K, Inoko H, Kulski JK. The HLA genomic loci map: expression, interaction, diversity and disease. J Hum Genet. 2009;54:15–39.
6. Johnson T, Barton N. Theoretical models of selection and mutation on quantitative traits. Philos Trans R Soc Lond B Biol Sci. 2005;360:1411–25.
7. Georgiou G, et al. The promise and challenge of high-throughput sequencing of the antibody repertoire. Nat Biotechnol. 2014;32:158–68.
8. Mitchell AM, Michels AW. T cell receptor sequencing in autoimmunity. J Life Sci Westlake Village Calif. 2020;2:38–58.
9. Mijnheer G, van Wijk F. T-cell compartmentalization and functional adaptation in autoimmune inflammation: lessons from pediatric rheumatic diseases. Front Immunol. 2019;10:940.
10. Petrelli A, van Wijk F. CD8(+) T cells in human autoimmune arthritis: the unusual suspects. Nat Rev Rheumatol. 2016;12:421–8.
11. Buniello A, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019;47:D1005–12.
12. López-Isac E. et al. Combined genetic analysis of juvenile idiopathic arthritis clinical subtypes identifies novel risk loci, target genes and key regulatory mechanisms. Ann Rheum Dis. 2020; annrheumdis-2020-218481. https://doi.org/10.1136/annrheumdis-2020-218481.
13. Bulik-Sullivan BK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47:291–5.
14. Aguet F, et al. Genetic effects on gene expression across human tissues. Nature. 2017;550:204–13.
15. Stancáková A, et al. Hyperglycemia and a common variant of GCKR are associated with the levels of eight amino acids in 9,369 Finnish men. Diabetes. 2012;61:1895–902.
16. Stancáková A, et al. Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men. Diabetes. 2009;58:1212–21.
17. Wright FA, et al. Heritability and genomics of gene expression in peripheral blood. Nat Genet. 2014;46:430–7.
18. Nuotio J, et al. Cardiovascular risk factors in 2011 and secular trends since 2007: the Cardiovascular Risk in Young Finns Study. Scand J Public Health. 2014;42:563–71.
19. Raitakari OT, et al. Cohort profile: the cardiovascular risk in Young Finns Study. Int J Epidemiol. 2008;37:1220–6.
20. 1000 Genomes Project Consortium, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491:56–65.
21. Mancuso N, et al. Large-scale transcriptome-wide association study identifies new prostate cancer risk regions. Nat Commun. 2018;9:4079.
22. Sun BB, et al. Genomic atlas of the human plasma proteome. Nature. 2018;558:73–9.
23. Zhang J, et al. Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies. Nat Genet. 2022;54:593–602.
24. Wang G, Sarkar A, Carbonetto P, Stephens M. A simple new approach to variable selection in regression, with application to genetic fine mapping. J R Stat Soc Ser B Stat Methodol. 2020;82:1273–300.
25. Mancuso N, et al. Probabilistic fine-mapping of transcriptome-wide association studies. Nat Genet. 2019;51:675–82.
26. Pain O, et al. Novel insight into the etiology of autism spectrum disorder gained by integrating expression data with genome-wide association statistics. Biol Psychiatry. 2019;86:265–73.
27. Liberzon A. et al. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011; 27: 1739–1740.
28. Pico AR, et al. WikiPathways: pathway editing for the people. PLOS Biol. 2008;6:e184.
29. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27–30.
30. Nishimura D. BioCarta. Biotech Softw Internet Rep. 2001;2:117–20.
31. Fabregat A, et al. Reactome pathway analysis: a high-performance in-memory approach. BMC Bioinformatics. 2017;18:142.
32. Schaefer CF, et al. PID: the Pathway Interaction Database. Nucleic Acids Res. 2009;37:D674-679.
33. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015;11:e1004219.
34. Altshuler DM, et al. Integrating common and rare genetic variation in diverse human populations. Nature. 2010;467:52–8.
35. Bulik-Sullivan B, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015;47:1236–41.
36. Zheng J, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics. 2017;33:272–9.
37. Onengut-Gumuscu S, et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. Nat Genet. 2015;47:381–6.
38. Schenck S, et al. Comorbidity of type 1 diabetes mellitus in patients with juvenile idiopathic arthritis. J Pediatr. 2018;192:196–203.
39. Shi H, Kichaev G, Pasaniuc B. Contrasting the genetic architecture of 30 complex traits from summary association data. Am J Hum Genet. 2016;99:139–53.
40. Barnes MG, et al. Subtype-specific peripheral blood gene expression profiles in recent-onset juvenile idiopathic arthritis. Arthritis Rheum. 2009;60:2102–12.
41. Mo A, et al. Disease-specific regulation of gene expression in a comparative analysis of juvenile idiopathic arthritis and inflammatory bowel disease. Genome Med. 2018;10:48.
42. Wong L, et al. Limits of Peripheral Blood Mononuclear Cells for Gene Expression-Based Biomarkers in Juvenile Idiopathic Arthritis. Sci Rep. 2016;6:29477.
43. Jung J, Kim GW, Lee B, Joo JWJ, Jang W. Integrative genomic and transcriptomic analysis of genetic markers in Dupuytren's disease. BMC Med Genomics. 2019;12:98.
44. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
45. Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? PLoS Comput Biol. 2011;7:e1001057.
46. Sergushichev, A. A. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. 2016. 060012

Kim *et al. Arthritis Research & Therapy*   (2024) 26:47

Page 18 of 19

https://www.biorxiv.org/content/10.1101/060012v1. https://doi.org/10.1101/060012.

47. Jiao X, et al. DAVID-WS: a stateful web service to facilitate gene/protein list analysis. Bioinforma Oxf Engl. 2012;28:1805–6.

48. Finucane HK, et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat Genet. 2018;50:621–9.

49. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015; 348: 648–660.

50. Pers TH, et al. Biological interpretation of genome-wide association studies using predicted gene functions. Nat Commun. 2015;6:5890.

51. Fehrmann RSN, et al. Gene expression analysis identifies global gene dosage sensitivity in cancer. Nat Genet. 2015;47:115–25.

52. Heng TSP, Painter MW & Immunological Genome Project Consortium. The Immunological Genome Project: networks of gene expression in immune cells. Nat. Immunol. 2008; 9: 1091–1094.

53. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015; 518: 317–330.

54. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57–74.

55. Corces MR, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat Genet. 2016;48:1193–203.

56. Berisa T, Pickrell JK. Approximately independent linkage disequilibrium blocks in human populations. Bioinforma Oxf Engl. 2016;32:283–5.

57. Yao DW, O'Connor LJ, Price AL, Gusev A. Quantifying genetic effects on disease mediated by assayed gene expression levels. Nat Genet. 2020;52:626–33.

58. Võsa U, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat Genet. 2021;53:1300–10.

59. Zeng J, et al. Widespread signatures of natural selection across human complex traits and functional genomic categories. Nat Commun. 2021;12:1164.

60. Boegel S, et al. HLA typing from RNA-Seq sequence reads. Genome Med. 2012;4:102.

61. Orenbuch R, et al. arcasHLA: high-resolution HLA typing from RNAseq. Bioinformatics. 2020;36:33–40.

62. Kim HJ, Pourmand N. HLA typing from RNA-seq data using hierarchical read weighting [corrected]. PloS One. 2013;8:e67885.

63. Nariai N, et al. HLA-VBSeq: accurate HLA typing at full resolution from whole-genome sequencing data. BMC Genomics. 2015;16:S7.

64. Szolek A, et al. OptiType: precision HLA typing from next-generation sequencing data. Bioinforma Oxf Engl. 2014;30:3310–6.

65. Bai Y, Wang D, Fury W. PHLAT: Inference of High-Resolution HLA Types from RNA and Whole Exome Sequencing. Methods Mol Biol Clifton NJ. 2018;1802:193–201.

66. Kawaguchi S, Higasa K, Shimizu M, Yamada R, Matsuda F. HLA-HD: An accurate HLA typing algorithm for next-generation sequencing data. Hum Mutat. 2017;38:788–97.

67. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. Bioinforma Oxf Engl. 2013;29:15–21.

68. Robinson J, et al. IPD-IMGT/HLA Database. Nucleic Acids Res. 2020;48:D948–55.

69. Robinson J, Soormally AR, Hayhurst JD, Marsh SGE. The IPD-IMGT/HLA Database - New developments in reporting HLA variation. Hum Immunol. 2016;77:233–7.

70. Mangul S, et al. ROP: dumpster diving in RNA-sequencing to find the source of 1 trillion reads across diverse adult human tissues. Genome Biol. 2018;19:36.

71. Mandric I, et al. Profiling immunoglobulin repertoires across multiple human tissues using RNA sequencing. Nat Commun. 2020;11:3126.

72. Maschmeyer P, et al. Antigen-driven PD-1+ TOX+ BHLHE40+ and PD-1+ TOX+ EOMES+ T lymphocytes regulate juvenile idiopathic arthritis in situ. Eur J Immunol. 2021;51:915–29.

73. Zheng GXY, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017;8:14049.

74. Hao Y, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184:3573-3587.e29.

75. Korsunsky I, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16:1289–96.

76. McInnes L, Healy J & Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. http://arxiv.org/abs/1802.03426 (2020). https://doi.org/10.48550/arXiv.1802.03426.

77. Wojdas M, Dąbkowska K, Winsz-Szczotka K. Alterations of extracellular matrix components in the course of juvenile idiopathic arthritis. Metabolites. 2021;11:132.

78. Jiang K, Sun X, Chen Y, Shen Y, Jarvis JN. RNA sequencing from human neutrophils reveals distinct transcriptional differences associated with chronic inflammatory states. BMC Med Genomics. 2015;8:55.

79. Hou X, et al. The multi-omics architecture of juvenile idiopathic arthritis. Cells. 2020;9:E2301.

80. Omoyinmi E, et al. Mitochondrial and oxidative stress genes are differentially expressed in neutrophils of sJIA patients treated with tocilizumab: a pilot microarray study. Pediatr Rheumatol Online J. 2016;14:7.

81. Liu S, et al. Integrative Analysis of Transcriptome-Wide Association Study and Gene-Based Association Analysis Identifies In Silico Candidate Genes Associated with Juvenile Idiopathic Arthritis. Int J Mol Sci. 2022;23:13555.

82. Mertz EL, et al. Matrix disruptions, growth, and degradation of cartilage with impaired sulfation. J Biol Chem. 2012;287:22030–42.

83. Hsu W, Rosenquist GL, Ansari AA, Gershwin ME. Autoimmunity and tyrosine sulfation. Autoimmun Rev. 2005;4:429–35.

84. Copland A, Bending D. Foxp3 molecular dynamics in treg in juvenile idiopathic arthritis. Front Immunol. 2018;9:2273.

85. Rubin CI, Atweh GF. The role of stathmin in the regulation of the cell cycle. J Cell Biochem. 2004;93:242–50.

86. Filbert EL, Le Borgne M, Lin J, Heuser JE, Shaw AS. Stathmin regulates microtubule dynamics and microtubule organizing center polarization in activated T cells. J Immunol Baltim Md. 2012;1950(188):5421–7.

87. Finucane HK, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet. 2015;47:1228–35.

88. Zeng J, et al. Signatures of negative selection in the genetic architecture of human complex traits. Nat Genet. 2018;50:746–53.

89. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet. 2001;69:124–37.

90. Eyre-Walker A. Evolution in health and medicine Sackler colloquium: Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. Proc Natl Acad Sci USA. 2010;107(Suppl 1):1752–6.

91. Gough SCL, Simmonds MJ. The HLA Region and Autoimmune Disease: Associations and Mechanisms of Action. Curr Genomics. 2007;8:453–65.

92. Cho JH, Feldman M. Heterogeneity of autoimmune diseases: pathophysiologic insights from genetics and implications for new therapies. Nat Med. 2015;21:730–8.

93. Li YR, et al. Genetic sharing and heritability of paediatric age of onset autoimmune diseases. Nat Commun. 2015;6:8442.

94. International HIV Controllers Study et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science. 2010; 330: 1551–1557.

95. Raychaudhuri S, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet. 2012;44:291–6.

96. Snyder A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. N Engl J Med. 2014;371:2189–99.

97. Evans DM, et al. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. Nat Genet. 2011;43:761–7.

98. Jones EY, Fugger L, Strominger JL, Siebold C. MHC class II proteins and disease: a structural perspective. Nat Rev Immunol. 2006;6:271–82.

99. Hinks A, et al. Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis. Nat Genet. 2013;45:664–9.

100. Ombrello MJ, et al. Genetic architecture distinguishes systemic juvenile idiopathic arthritis from other forms of juvenile idiopathic arthritis: clinical and therapeutic implications. Ann Rheum Dis. 2017;76:906–13.

101. Friese MA, Jones EY, Fugger L. MHC II molecules in inflammatory diseases: interplay of qualities and quantities. Trends Immunol. 2005;26:559–61.

Kim *et al. Arthritis Research & Therapy*   (2024) 26:47

Page 19 of 19

102. Luo Y, et al. A high-resolution HLA reference panel capturing global population diversity enables multi-ancestry fine-mapping in HIV host response. Nat Genet. 2021;53:1504–16.

103. Bréhant J. [The centennial of appendicectomy. Apropos of the report of B. Godquin]. Chir. Memoires Acad. Chir. 1987; 113: 753.

104. Shale M, Schiering C, Powrie F. CD4($+$) T-cell subsets in intestinal inflammation. Immunol Rev. 2013;252:164–82.

105. Van Boxel JA, Paget SA. Predominantly T-cell infiltrate in rheumatoid synovial membranes. N Engl J Med. 1975;293:517–20.

106. Wedderburn LR, Robinson N, Patel A, Varsani H, Woo P. Selective recruitment of polarized T cells expressing CCR5 and CXCR3 to the inflamed joints of children with juvenile idiopathic arthritis. Arthritis Rheum. 2000;43:765–74.

107. Barennes P, et al. Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases. Nat Biotechnol. 2021;39:236–45.

108. De Simone M, Rossetti G, Pagani M. Single Cell T Cell Receptor Sequencing: Techniques and Future Challenges. Front Immunol. 2018;9:1638.

109. Schwartz AB, et al. Elucidating the Biomechanics of Leukocyte Transendothelial Migration by Quantitative Imaging. Front Cell Dev Biol. 2021;9:635263.

110. Zhong J, et al. A transcriptome-wide association study identifies novel candidate susceptibility genes for pancreatic cancer. JNCI. 2020. https://doi.org/10.1093/jnci/djz246.

111. Pérès E, et al. PDZ domain-binding motif of Tax sustains T-cell proliferation in HTLV-1-infected humanized mice. PLoS Pathog. 2018;14:e1006933.

112. Norén E, Almer S, Söderman J. Genetic variation and expression levels of tight junction genes identifies association between MAGI3 and inflammatory bowel disease. BMC Gastroenterol. 2017;17:68.

113. Picard C, et al. Primary Immunodeficiency Diseases: an Update on the Classification from the International Union of Immunological Societies Expert Committee for Primary Immunodeficiency 2015. J Clin Immunol. 2015;35:696–726.

114. Hashiguchi K, Ozaki M, Kuraoka I, Saitoh H. Establishment of a human cell line stably overexpressing mouse Nip45 and characterization of Nip45 subcellular localization. Biochem Biophys Res Commun. 2013;430:72–7.

115. Yoshimura T, et al. Two-sided roles of IL-27: induction of Th1 differentiation on naive CD4$+$ T cells versus suppression of proinflammatory cytokine production including IL-23-induced IL-17 on activated CD4$+$ T cells partially through STAT3-dependent mechanism. J Immunol Baltim Md. 2006;1950(177):5377–85.

116. Gaur P, Misra R, Aggarwal A. IL-27 levels are low in enthesitis-related arthritis category of juvenile idiopathic arthritis. Clin Exp Rheumatol. 2016;34:337–42.

117. Babaie F, et al. The roles of ERAP1 and ERAP2 in autoimmunity and cancer immunity: New insights and perspective. Mol Immunol. 2020;121:7–19.

118. Zaripova LN, et al. Juvenile idiopathic arthritis: from aetiopathogenesis to therapeutic approaches. Pediatr Rheumatol Online J. 2021;19:135.

119. Feng R, et al. Identification of candidate genes and pathways associated with juvenile idiopathic arthritis by integrative transcriptome-wide association studies and mRNA expression profiles. Arthritis Res Ther. 2023;25:19.

120. Xu J, et al. A cross-tissue transcriptome-wide association study identifies novel susceptibility genes for juvenile idiopathic arthritis in Asia and Europe. Front Immunol. 2022;13:941398.

121. Liu X, Li YI, Pritchard JK. Trans effects on gene expression can drive omnigenic inheritance. Cell. 2019;177:1022-1034.e6.

122. Sinnott-Armstrong N, Naqvi S, Rivas M & Pritchard JK. GWAS of three molecular traits highlights core genes and pathways alongside a highly polygenic background. eLife. 2021; 10: e58615.

123. Hersh AO, Prahalad S. Immunogenetics of juvenile idiopathic arthritis: a comprehensive review. J Autoimmun. 2015;64:113–24.

124. Cao K, et al. Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations. Hum Immunol. 2001;62:1009–30.

125. Janse van Rensburg WJ, de Kock A, Bester C & Kloppers JF. HLA major allele group frequencies in a diverse population of the Free State Province, South Africa. Heliyon. 2021; 7: e06850.

126. De Silvestri A, et al. HLA-DRB1 alleles and juvenile idiopathic arthritis: diagnostic clues emerging from a meta-analysis. Autoimmun Rev. 2017;16:1230–6.

127. Matzaraki V, Kumar V, Wijmenga C, Zhernakova A. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. Genome Biol. 2017;18:76.

128. Yanagimachi M, et al. Association of HLA-A*02:06 and HLA-DRB1*04:05 with clinical subtypes of juvenile idiopathic arthritis. J Hum Genet. 2011;56:196–9.

129. Lee YH, Bae SC & Song GG. The association between the functional PTPN22 1858 C/T and MIF -173 C/G polymorphisms and juvenile idiopathic arthritis: a meta-analysis. Inflamm Res Off J Eur Histamine Res Soc. 2012; AI 61: 411–415.

130. Ishigaki K, et al. HLA autoimmune risk alleles restrict the hypervariable region of T cell receptors. Nat Genet. 2022;54:393–402.

131. Viret C, Janeway CA. MHC and T cell development. Rev Immunogenet. 1999;1:91–104.

132. Wucherpfennig KW, Sethi D. T cell receptor recognition of self and foreign antigens in the induction of autoimmunity. Semin Immunol. 2011;23:84–91.

133. Spreafico R, et al. A circulating reservoir of pathogenic-like CD4+ T cells shares a genetic and phenotypic signature with the inflamed synovial micro-environment. Ann Rheum Dis. 2016;75:459–65.

134. Analysis of T Cell Repertoire Diversity of CD4+ Memory and NaïVe T Cells By Next Generation Sequencing and Its Association with Rheumatoid Arthritis Disease Parameters. ACR Meeting Abstracts https://acrabstracts.org/abstract/analysis-of-t-cell-repertoire-diversity-of-cd4-memory-and-naive-t-cells-by-next-generation-sequencing-and-its-association-with-rheumatoid-arthritis-disease-parameters/.

135. Liu X, et al. T cell receptor β repertoires as novel diagnostic markers for systemic lupus erythematosus and rheumatoid arthritis. Ann Rheum Dis. 2019;78:1070–8.

136. Wagner UG, Koetz K, Weyand CM, Goronzy JJ. Perturbation of the T cell repertoire in rheumatoid arthritis. Proc Natl Acad Sci U S A. 1998;95:14447–52.

137. Kobayashi S, et al. A distinct human CD4+ T cell subset that secretes CXCL13 in rheumatoid synovium. Arthritis Rheum. 2013;65:3063–72.

138. Peterson RE, et al. Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. Cell. 2019;179:589–603.

139. Chen X, et al. Racial/Ethnic Differences in Sleep Disturbances: The Multi-Ethnic Study of Atherosclerosis (MESA). Sleep. 2015;38:877–88.

Publisher's Note