




# Genetic variants shape rheumatoid arthritis-specific transcriptomic features in CD4<sup>+</sup> T cells through differential DNA methylation, explaining a substantial proportion of heritability

Eunji Ha <sup>1</sup>, So-Young Bang,<sup>2,3</sup> Jiwoo Lim,<sup>1</sup> Jun Ho Yun,<sup>4</sup> Jeong-Min Kim,<sup>4</sup> Jae-Bum Bae,<sup>4</sup> Hye-Soon Lee,<sup>2,3</sup> Bong-Jo Kim,<sup>4</sup> Kwangwoo Kim <sup>1</sup>, Sang-Cheol Bae <sup>2,3</sup>

**Handling editor** Josef S Smolen

► Additional material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/annrheumdis-2020-219152>).

For numbered affiliations see end of article.

## Correspondence to

Professor Sang-Cheol Bae, Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, Republic of Korea; scbae@hanyang.ac.kr, Professor Kwangwoo Kim, Department of Biology, Kyung Hee University, Seoul, Republic of Korea; kkim@khu.ac.kr and Dr Bong-Jo Kim, Division of Genome Science, Department of Precision Medicine, National Institute of Health, Chungcheongbuk-do, Republic of Korea; kbj6181@korea.kr

EH and S-YB contributed equally.

Received 21 September 2020  
Revised 10 December 2020  
Accepted 30 December 2020  
Published Online First  
12 January 2021



© Author(s) (or their employer(s)) 2021. No commercial re-use. See rights and permissions. Published by BMJ.

**To cite:** Ha E, Bang S-Y, Lim J, et al. *Ann Rheum Dis* 2021;**80**:876–883.

## ABSTRACT

**Objective** CD4<sup>+</sup> T cells have been suggested as the most disease-relevant cell type in rheumatoid arthritis (RA) in which RA-risk non-coding variants exhibit allele-specific effects on regulation of RA-driving genes. This study aimed to understand RA-specific signatures in CD4<sup>+</sup> T cells using multi-omics data, interpreting inter-omics relationships in shaping the RA transcriptomic landscape.

**Methods** We profiled genome-wide variants, gene expression and DNA methylation in CD4<sup>+</sup> T cells from 82 patients with RA and 40 healthy controls using high-throughput technologies. We investigated differentially expressed genes (DEGs) and differential methylated regions (DMRs) in RA and localised quantitative trait loci (QTLs) for expression and methylation. We then integrated these based on individual-level correlations to inspect DEG-regulating sources and investigated the potential regulatory roles of RA-risk variants by a partitioned-heritability enrichment analysis with RA genome-wide association summary statistics.

**Results** A large number of RA-specific DEGs were identified (n=2575), highlighting T cell differentiation and activation pathways. RA-specific DMRs, preferentially located in T cell regulatory regions, were correlated with the expression levels of 548 DEGs mostly in the same topologically associating domains. In addition, expressional variances in 771 and 83 DEGs were partially explained by expression QTLs for DEGs and methylation QTLs (meQTLs) for DEG-correlated DMRs, respectively. A large number of RA variants were moderately to strongly correlated with meQTLs. DEG-correlated DMRs, enriched with meQTLs, had strongly enriched heritability of RA.

**Conclusion** Our findings revealed that the methylomic changes, driven by RA heritability-explaining variants, shape the differential expression of a substantial fraction of DEGs in CD4<sup>+</sup> T cells in patients with RA, reinforcing the importance of a multidimensional approach in disease-relevant tissues.

## INTRODUCTION

Rheumatoid arthritis (RA) is an inflammatory autoimmune disease causing chronic symmetrical polyarthritis of large and small joints and mostly occurs in women between 30 and 50 years of age.<sup>1</sup> The causes of RA are not yet fully understood

## Key messages

### What is already known about this subject?

- Rheumatoid arthritis (RA) has a highly polygenic genetic architecture, with nearly 120 reported RA susceptibility loci and a large number of unidentified RA loci.
- The majority of disease variants are located in non-coding elements, most of which significantly overlap with CD4<sup>+</sup> T cell regulatory elements. In addition, many genes in RA loci are involved in CD4<sup>+</sup> T cell pathways.

### What does this study add?

- This study provides the landscapes of transcriptomic and methylomic features in RA CD4<sup>+</sup> T cells, with catalogues of quantitative trait loci for expression and methylation.
- The integrative approaches using individual-level genetic, epigenetic and transcriptomic data with recent Korean genome-wide RA association statistics dissected the regulatory sources for differentially expressed genes in RA CD4<sup>+</sup> T cells, newly suggesting that the RA-risk variant-driven methylation changes result in the differential expression of a large number of the genes in RA CD4<sup>+</sup> T cells.

### How might this impact on clinical practice or future developments?

- Our findings contribute to a better understanding of the CD4<sup>+</sup> T cell alterations underlying RA association of non-coding variants and identify disease-relevant gene candidates that may be used for novel therapeutic targets.

but a family-based genetic approach estimated an overall genetic heritability of RA to be up to 65%.<sup>2</sup> Genome-wide association studies (GWASs) revealed a highly polygenic genetic aetiology of RA, identifying RA-associated common variants in ~120 susceptibility loci.<sup>3–5</sup> Most (~90%) of the RA-risk association signals in RA loci come from non-coding variants. The biological functions of most non-coding RA-risk variants have been largely

unknown,<sup>6</sup> but bioinformatic analyses using non-coding RA-risk variants and cell type-specific features (eg, regulatory annotations and cellular pathways) were highly fruitful in narrowing down disease-relevant cell types in RA pathogenesis.

Several studies reported that the non-coding variants, which were associated with autoimmune disorders, were significantly enriched within enhancers and around expression quantitative trait loci (eQTLs) in immune cell types,<sup>7–10</sup> implying a possible allele-specific regulatory effect of non-coding variants in disease-relevant cell types in the pathogenesis of autoimmune disorders. Several studies on the RA-risk non-coding variants highlighted the importance of CD4<sup>+</sup> T cells in RA. The RA-risk variants are strongly enriched in cell type-specific annotations including enhancers, histone modification marks and transcription factor-binding sites (TFBSs) in CD4<sup>+</sup> T cells in RA pathogenesis.<sup>5 7 9–12</sup> Indeed, a large number of the genes within RA susceptibility loci are involved in the activation and differentiation pathways of CD4<sup>+</sup> T cells.<sup>3 13–15</sup>

Despite the significant enrichment of non-coding disease variants in cell-type specific annotations, only a minor fraction (10% to 20%) of the GWAS signals in autoimmune diseases including RA were directly explained by known eQTLs or TFBSs.<sup>6 16</sup> Therefore, a new approach is needed to understand how the disease variants exert regulatory effects on disease effector genes. Indirect regulatory effects of disease variants through epigenetic changes are likely to be undetectable in limited sample sizes of most eQTL analyses.

Here, we generated genomic, transcriptomic and epigenomic (DNA methylation) data from purified CD4<sup>+</sup> T cells in identical patients with RA and healthy controls. This study provides comprehensive landscapes of RA-specific transcriptomic and epigenomic signatures in CD4<sup>+</sup> T cells, identifies the variants associated with expression or methylation levels and integrates them with recent Korean GWAS data<sup>4</sup> to understand how RA heritability-explaining variants shape RA-specific differential expression in CD4<sup>+</sup> T cells on a genome-wide scale.

## METHODS

### Subjects and CD4<sup>+</sup> T cell isolation

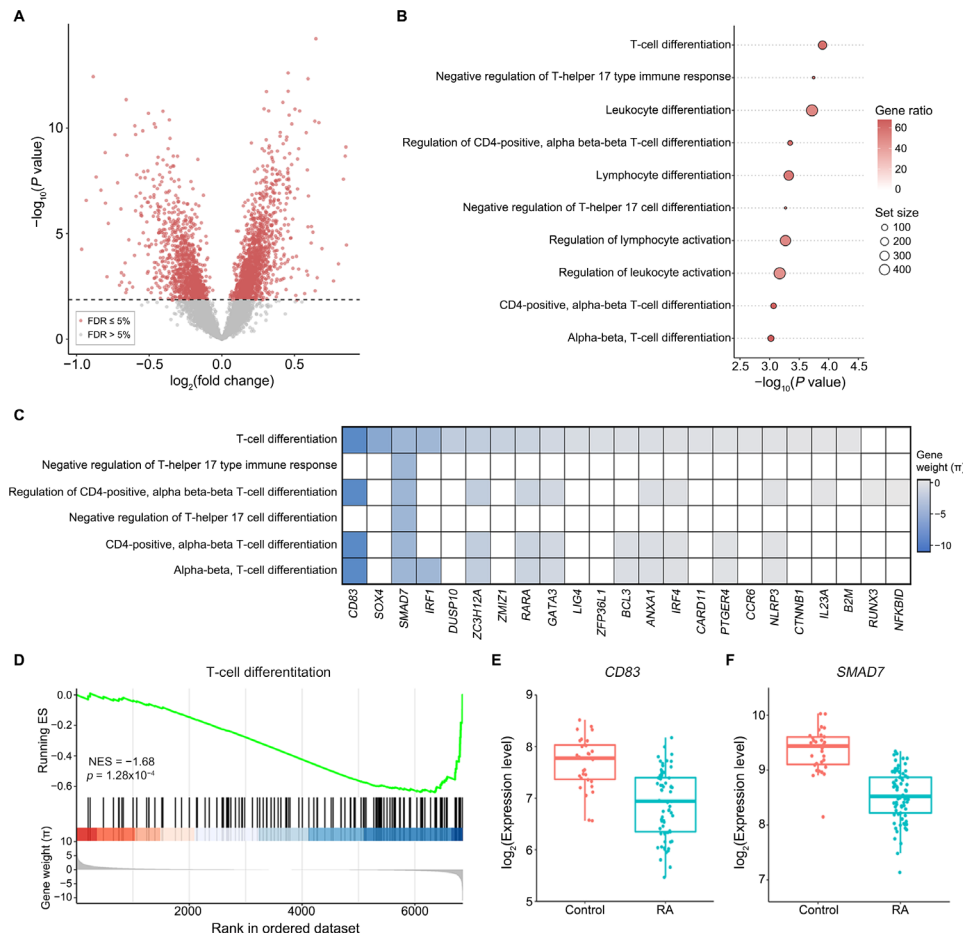
A total of 122 study subjects consisting of 82 patients with RA and 40 healthy controls were recruited at Hanyang University Hospital for Rheumatic Diseases (Seoul, South Korea). Sample sizes in each omics data set are summarised in figure 1. All the subjects provided written informed consent for participation. After collecting ~16 cc peripheral blood mononuclear cells using BD Vacutainer CPT, CD4<sup>+</sup> T cells were purified within 3 hour (without freezing and thawing) using Invitrogen Dynabeads CD4<sup>+</sup> Isolation Kit, and genomic DNA and messenger RNA were extracted. The CD4<sup>+</sup> T cell purification method was internally verified to yield a high purity of CD4<sup>+</sup> T cells by fluorescence-activated cell sorting and all samples showed >90% purity of CD4<sup>+</sup> T cells in a methylation-based cell composition analysis (online supplemental figure S1). All patients with RA were diagnosed according to the American College of Rheumatology 1987 classification criteria for RA.<sup>17</sup> The characteristics of the study subjects at the time of blood sampling are provided in online supplemental table S1.

### Gene expression analysis

The gene expression level in CD4<sup>+</sup> T cells was measured using the Illumina HumanHT-12 v4 BeadChip. Normal exponential background correction and quantile normalisation were performed for each slide using limma.<sup>18</sup> A total of 9414 expression probes were retained after a general quality control (QC) procedure (see the details in online supplemental table S2). A batch effect was removed by ComBat implemented in sva<sup>19</sup> using batch variables (slide and array position; online supplemental figure S2). A multivariate linear model was applied using limma to identify differentially expressed genes (DEGs) in RA at a false discovery rate (FDR) threshold of 0.05, controlling for potential confounders (sex, age and T cell purity) and computing moderated statistics by empirical Bayes shrinkage.<sup>18</sup>

		Methylome (n=122)	Transcriptome (n=103)	Genome (n=104)
		Methylation 450K BeadChip (n=122) or MBD-Seq (n=68)	HumanHT-12 v4 BeadChip	Genome-wide SNP array (KoreanChip)
	Cases	Controls		
Sample size (n)	64	26	✓	✓
	9	4	✓	
	7	7	✓	✓
	2	3	✓	

**Figure 1** Summary of sample sizes and overlap in three omics data sets. The study subjects consisted of 82 patients with rheumatoid arthritis and 40 healthy controls. Methylome data were generated for all the subjects (n=122) using a methylation array; a subset of the same subjects was used to generate MBD-Seq methylome data. Transcriptome and genome data for subsets (n=103 and 104, respectively) were generated using array technologies. The sample sizes were additionally summarised according to the combination of available omics data sets. MBD-Seq, methyl-CpG-binding domain sequencing.



**Figure 2** DEGs in RA CD4<sup>+</sup> T cells and DEG-enriched gene sets. (A) The statistical significance level for differential expression (y-axis; a negative  $\log_{10}$  scale) was plotted according to the  $\log_2$ -transformed fold change in expression in patients with RA compared with controls (x-axis). Significant data points above a significance threshold (depicted as a dashed line; FDR of 5%) are marked in red. (B) Dot plot represents the GO terms significantly enriching DEGs in a GSEA. The significance of enrichment is shown on the x-axis. The gene ratio indicates the ratio of the number of input genes to the total number of genes in the gene set (=set size). (C) Heat map shows the  $\pi$  values of the genes strongly contributing to the enrichment in T cell pathways. (D) Enrichment plot was created for the GO term, T cell differentiation with normalised enrichment score (NES)=-1.68 and enrichment p value= $1.28 \times 10^{-4}$ . The x-axis represents query genes ordered by their  $\pi$  values (shown in the lower part of the plot). The upper part of the plot shows the running enrichment score (ES) that means the sum of ESs from the top ranked gene to a corresponding gene. Bars in the middle part indicate the location of gene members of the query GO term. (E and F) Box plots represent normalised expression levels of (E) *CD83* and (F) *SMAD7* on a  $\log_2$  scale. The expressions of both genes were decreased in patients with RA compared with controls (fold change=-0.89 and  $P_{\text{FDR}}=1.1 \times 10^{-9}$  for *CD83*; fold change=-0.83 and  $P_{\text{FDR}}=6.0 \times 10^{-6}$  for *SMAD7*). DEGs, differentially expressed genes; FDR, false discovery rate; GSEA, gene set enrichment analysis;  $P_{\text{FDR}}$ , FDR-corrected p values; RA, rheumatoid arthritis.

### Gene set enrichment analysis

For each gene tested in the DEG analysis,  $\pi$ -value<sup>20</sup> was calculated based on the  $\log_2$ -fold expression change (*K*) in patients with RA and its significance level (*p*), as follows:

$$\pi = K \cdot (-\log_{10} p)$$

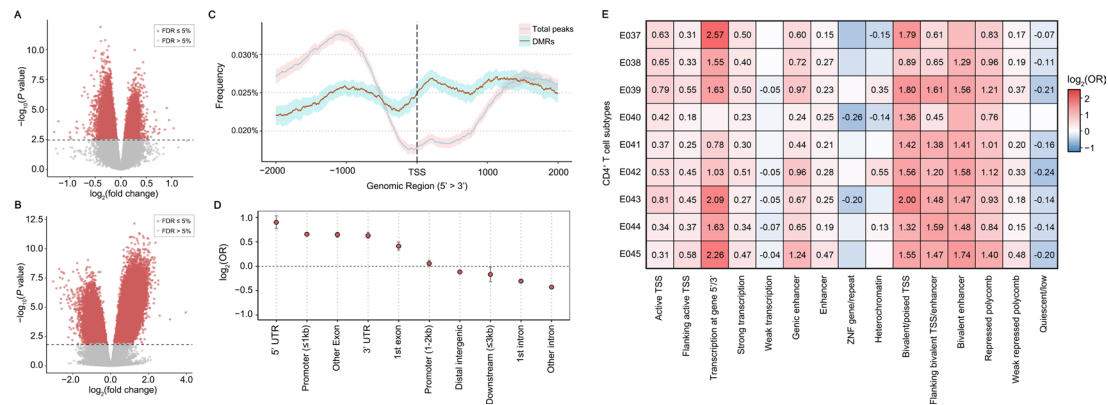
If a gene was analysed by multiple probes, the probe with the lowest *p* value for the differential expression was used in  $\pi$  calculation. A  $\pi$ -based gene set enrichment analysis (GSEA) for immune-related biological process terms under GO:0002376 was performed using clusterProfiler.<sup>21</sup>

### DNA methylation analysis

DNA methylation data were generated from genomic DNAs in CD4<sup>+</sup> T cells using both methyl-CpG-binding domain sequencing (MBD-Seq) for profiling of DNA methylation on a whole-genome scale and the Illumina Infinium Human Methylation 450K BeadChip mostly for profiling of CpG sites

around genic regions. In the MBD-Seq data analysis, reads were mapped to the human reference genome hg38 using bowtie2<sup>22</sup> and filtered out with <10 MAPQ values. Methylation peaks were called by MACS2<sup>23</sup> with default parameters and quantified by DiffBind<sup>24</sup> in 304 301 regions found in >1/3 in each group. Differentially methylated peaks (DMPeaks) were examined by DESeq2<sup>25</sup> with the same confounding factors used in the DEG analysis.

In methylation array data analysis, 413 718 CpG-targeting probes passed general QC filters (see the details in online supplemental table S3) and were analysed to identify differentially methylated probes (DMProbes) in RA using ChAMP.<sup>26</sup> Specifically, the fluorescence intensities of array probes were normalised by beta-mixture quantile normalisation<sup>27</sup> and transformed into M value<sup>28</sup> to avoid heteroscedasticity. We eliminated the batch effect using batch variables (data production time, slide and array position) by ComBat implemented in *sva*<sup>19</sup> (online supplemental figure S3). DMProbes between patients



**Figure 3** DMRs and DMR-enriched genomic annotations in RA CD4<sup>+</sup> T cells. (A and B) Volcano plots were generated from the DMR analysis results using (A) methylation array and (B) MBD-Seq data. The negative log<sub>10</sub>-transformed statistical significance level for differential methylation (y-axis) was plotted according to log<sub>2</sub>-transformed fold change in methylation in patients with RA compared with controls (x-axis). Significant data points (DMProbes and DMPeaks) above a significance threshold (depicted as a dashed line; FDR of 5%) are marked in red. (C) Distribution of DMPeaks and total MBD-Seq peaks is shown based on the distance from TSSs. The 95% CIs estimated by bootstrapping are shown as shaded areas. (D) ORs for DMRs that are located in each genomic feature are shown with error bars indicating 95% CIs. (E) LOLA analysis results are summarised in the heat map highlighting T cell-specific ChromHMM chromatin states enriched with DMRs. Significant OR values (FDR ≤5%) are shown in the heat map. The Roadmap Epigenomics Project data was used in the analysis (E037, CD4<sup>+</sup> memory T cells; E038, CD4<sup>+</sup> naïve T cells; E039, CD4<sup>+</sup>CD25<sup>+</sup>CD45RA<sup>+</sup> T cells; E040, CD4<sup>+</sup>CD25<sup>+</sup>CD45RO<sup>+</sup> T cells; E041, stimulated CD4<sup>+</sup>CD25<sup>+</sup>IL17<sup>+</sup> T cells; E042, stimulated CD4<sup>+</sup>CD25<sup>+</sup>IL17<sup>+</sup> T cells; E043, CD4<sup>+</sup>CD25<sup>-</sup> cells; E044, CD4<sup>+</sup>CD25<sup>-</sup>IL127<sup>+</sup> T<sub>reg</sub> cells; E045, CD4<sup>+</sup>CD25<sup>-</sup>IL127<sup>+</sup> T cells). DMPeaks, differentially methylated peaks; DMProbes, differentially methylated probes; DMRs, differentially methylated regions; FDR, false discovery rate; MBD-Seq, methyl-CpG-binding domain sequencing; RA, rheumatoid arthritis; TSSs, transcription start sites.

with RA and controls were investigated under a multivariate linear model considering the same covariates used in the DEG analysis.

Differential methylation regions (DMRs; DMPeaks or DMProbes) with FDR-corrected p values ( $P_{\text{FDR}} < 0.05$ ) were considered as significant.

### Profiling the genome-wide landscape of RA-specific DMPeaks

Genome-wide DNA methylation data from MBD-Seq were used to profile the overlap between genomic annotations and DMPeaks in RA. Genomic annotations for CD4<sup>+</sup> T cell-specific ChromHMM chromatin states,<sup>29</sup> gene-based positional annotations and TFBSs were retrieved from the Roadmap Epigenomics Project data<sup>30</sup> and ChIPseeker<sup>31</sup> and PAINTOR ([https://github.com/gkichaev/PAINTOR\\_V3.0/wiki/2b.-Overlapping-annotations](https://github.com/gkichaev/PAINTOR_V3.0/wiki/2b.-Overlapping-annotations)), respectively. The distribution of DMPeaks around transcription start sites (TSSs) was drawn using ChIPseeker.<sup>31</sup> A Fisher's exact test was performed to assess an enrichment of DMPeaks on a query annotation using LOLA<sup>32</sup> with all methylation peaks as background.

### Correlation analysis between expression level of DEGs and methylation level of DMR

The RA-specific DEGs and DMRs in the same topologically associating domain (TAD) of CD4<sup>+</sup> T cells<sup>33</sup> (available at <https://osf.io/u8tzip>) were tested for their individual-level correlation. In the region not characterised by any TADs, DMRs in a 2 kb region around the TSS of a DEG were used in the pairwise correlation analysis. Significant individual-level correlations between DMR and DEGs were identified in a linear regression controlling for the same covariates used in the DEG analysis, at a gene-level FDR of 5%.

### Analysis of genetic associations with the level of expression and methylation

Linear regression was performed to identify *cis*-QTLs adjusting for sex, age, T cell composition, disease status and the top five genotypic principal components using FastQTL<sup>34</sup> at peak-wise or probe-wise FDR of 5%. *Cis*-variants within 1 Mb of TSSs or methylated regions were used in the analyses. We employed

QTLtools<sup>35</sup> to identify trans-QTLs ≥5 Mb away from each target site using the same model in *cis*-QTL analyses (p value threshold =  $1 \times 10^{-10}$ ). We retained a subset of independent QTLs by linkage disequilibrium (LD) clumping ( $r^2 > 0.2$ ) in each gene or methylated region.

The methylation-mediated effects of methylation QTLs (meQTLs) on regulation of DEGs were assessed by a mediation analysis with quasi-Bayesian CIs<sup>36</sup> at a gene-wide FDR of 5%.

### Estimating heritability of RA partitioned by RA-specific DEGs or DMRs

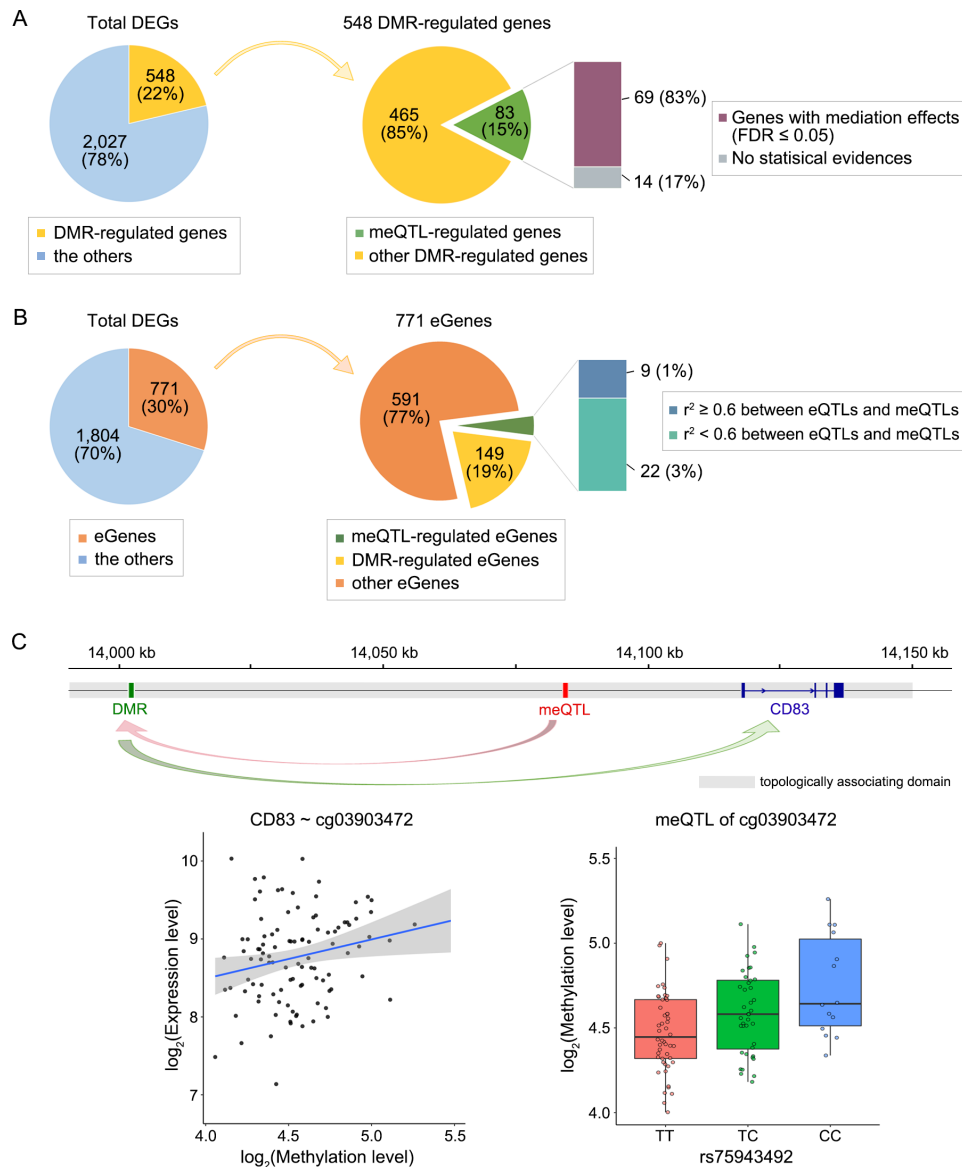
The enrichment of heritability ( $h^2$ ) of all variants in the identified annotations was estimated using stratified LD score regression<sup>7</sup> based on the 1KGP East Asian LD scores (<https://data.broadinstitute.org/alkesgroup/LDSCORE/>) and the RA association summary statistics of our recent Korean GWAS.<sup>4</sup> According to the observed distribution of eQTLs and meQTLs in this study, the regulatory genomic region enriched with eQTLs for DEGs or non-DEGs was defined as a 5 kb region around the TSS of each gene. Similarly, the genomic region enriched with meQTLs for DEG-expression-correlated DMRs, expression-uncorrelated DMRs or non-DMRs was defined as a methylation region of interest with a 5 kb buffer region.

Detailed methods of MBD-Seq data analysis, genotyping and whole-genome imputation, validation of eQTLs with public eQTLs, enrichment of QTLs on TFBSs and colocalisation test for RA association and QTL signals are described in online supplemental note.

## RESULTS

### Brief overview of the main analyses

This study consisted of four main analyses to understand RA-specific features in CD4<sup>+</sup> T cells at the level of genomics, methylomics and transcriptomics in a single cohort and to scope out the inter-omics relationship in regulating RA-specific DEGs in CD4<sup>+</sup> T cells. (1) An expression microarray analysis followed



**Figure 4** Summary of regulatory features associated with differential expression in RA CD4<sup>+</sup> T cells. (A and B) The numbers of DEGs are shown according to potentially regulatory features; DMRs, *cis*-meQTLs or *cis*-eQTLs. (A) DMR-regulated genes refer to genes whose expression level was significantly correlated with the methylation level of nearby DMRs in the same topologically associating domain (TAD) or the promoter region. If *cis*-meQTLs were detected in the corresponding DMRs, the genes are referred as meQTL-regulated genes. The mediation effects of *cis*-meQTLs on DEGs were estimated using a mediation analysis with quasi-Bayesian CIs. (B) Genes that were regulated by *cis*-eQTLs in this study or other studies were defined as eGenes. The eGenes were stratified into meQTL-regulated and DMR-regulated eGenes according to methylation features associated with eGenes. (C) The expression level of *CD83* was regulated by the meQTL rs75943492, mediated by the differential methylation on the DMP probe cg03903472 (mediation effect = -0.12,  $P_{FDR} = 0.01$ ) in a TAD. The associations were significant between expression and methylation levels ( $\beta = 0.55$ ,  $P_{FDR} = 0.03$ ) and between rs75943492 and methylation level ( $\beta = 0.16$ ,  $P_{FDR} = 0.02$ ). DEGs, differentially expressed genes; DMP probe, differentially methylated probe; DMRs, differentially methylated regions; eQTLs, expression quantitative trait loci; FDR, false discovery rate; meQTLs, methylation QTLs;  $P_{FDR}$ , FDR-corrected p-values; RA, rheumatoid arthritis.

by a GSEA was carried out to identify DEGs and DEG-enriched pathways in CD4<sup>+</sup> T cells in patients with RA. (2) DNA methylation analyses using both MBD-Seq and methylation arrays were conducted to determine genome-wide methylation profiles in CD4<sup>+</sup> T cells and RA-specific DMRs. (3) QTL analyses were performed by integrating genome-wide variant data with expression and methylation data in CD4<sup>+</sup> T cells. (4) An inter-omics analysis was performed to better understand plausible DEG-regulating mechanisms in RA CD4<sup>+</sup> T cells and the potential contribution of RA heritability-explaining variants.

### RA-specific DEGs in CD4<sup>+</sup> T cells highlighting the differentiation and activation of CD4<sup>+</sup> T cells

Expression analyses identified 2575 DEGs in RA CD4<sup>+</sup> T cells, based on the differential fluorescence signal of 2785 expression probes. Among the identified DEGs, 1585 and 1200 DEGs were upregulated and downregulated in RA, respectively (figure 2A). In a GSEA, RA-specific DEGs were significantly enriched in 10 immune processes, most of which were T cell-related pathways including lymphocyte activation and CD4<sup>+</sup> T cell differentiation (figure 2B–D and online supplemental table S5). The

most significant DEGs included some known genes relevant to T-cell biology (eg, *CD83*,<sup>37 38</sup> *SMAD7*<sup>39</sup> and *IRF1*<sup>40</sup>; figure 2C, E and F). For example, the anti-inflammatory gene *CD83*<sup>37</sup> in an RA susceptibility locus showed an approximate twofold decrease in expression level in patients with RA compared with controls ( $\log_2$  fold change = -0.89;  $P_{\text{FDR}} = 1.1 \times 10^{-9}$ ; figure 2E) and belonged to most of the identified pathways including T-cell differentiation, contributing to the pathway enrichment with the largest size of  $\pi$ -value<sup>20</sup> ( $\pi = -11.0$ ).

### RA-specific DMRs preferentially within regulatory regions

We used MBD-Seq and methylation microarrays to capture the genome-wide methylation architecture in CD4<sup>+</sup> T cells in RA, with a high resolution on genic CpG sites. We observed a high correlation between 300bp-bin MBD-Seq read counts and methylation probe intensities (Pearson's  $r = 0.75$  for cases, 0.71 for controls, (online supplemental figure S4). A total of 94 898 DMPeaks (30.4%; out of 304 301 peaks) and 28 786 DMProbes (7.0%; out of 413 718 probes) in RA were identified in MBD-Seq and microarray data, respectively (figure 3A,B). There was a significant overlap between DMPeaks and DMProbes ( $p = 9.9 \times 10^{-94}$  in a Fisher's exact test for the region tested in both analyses), showing the concordant direction of methylation changes in >90% of co-localising DMPeak-DMProbe pairs.

The unbiased genome-wide methylation landscape through MBD-Seq revealed how the DNA methylation sites are distributed based on genomic annotation and emphasised the strong enrichment of RA-specific DMRs in the likely regulatory region around TSSs (figure 3C), including the 5' UTR (OR = 1.87,  $p = 1.34 \times 10^{-39}$ ) and proximal promoters (OR = 1.58,  $p = 2.29 \times 10^{-245}$ ; figure 3D and online supplemental table S6). Consistently, DMRs were significantly enriched in CD4<sup>+</sup> T cell-specific ChromHMM chromatin states<sup>30</sup> associated with transcription-activating, repressing or bivalent regions (figure 3E).

### Identification of QTLs

We identified 2125 *cis*-eQTLs for 682 expression probes, 120 424 meQTLs for 43 526 methylation probes and 23 690 *cis*-meQTLs for 11 998 methylation peaks within 1 Mb of corresponding TSSs or methylated regions. The detected *cis*-eQTLs and *cis*-meQTLs were very closely located to corresponding TSSs and methylated regions, respectively (online supplemental figure S5). Although the sample size and data type<sup>41</sup> might be insufficient to ensure statistical power to detect QTLs, we observed that the identified CD4<sup>+</sup> T cell *cis*-eQTLs were highly consistent with publicly available *cis*-eQTLs from the CD4<sup>+</sup> T cell RNA-Seq data (online supplemental figure S6). Similarly, by comparing with the publicly available lead meQTLs in whole blood,<sup>42</sup> we found consistent effect sizes in our study (online supplemental figure S7).

We further identified trans-eQTLs for 17 expression probes, 233 trans-meQTLs for 234 methylation probes and 21 trans-meQTLs for 21 methylation peaks outside 5 Mb of the TSSs or methylated sites.

RA variants in many susceptibility loci are correlated moderately to strongly with eQTLs or meQTLs (online supplemental table S7-9). For example, the *cis*-eQTL signals (rs8046707) of *FBS* downregulated in RA were statistically showed  $r^2 = 0.42$  with a lead RA variant rs12918327 in the same locus (posterior probability of a shared causal signal >50% in a colocalisation analysis).

In addition, we observed that QTLs were significantly enriched on binding sites of 109 transcription factors. Most of the identified transcription factors also significantly bound within DMRs at FDR of 5% (online supplemental table S10 and figure S8). Some top-ranked transcription factors that preferentially bind to QTLs and DMRs are known to be relevant to RA or T cell functions (eg, MAZ, a Myc-associated protein.<sup>43</sup>)

### Methylation-mediation effects of meQTLs on RA-specific DEGs

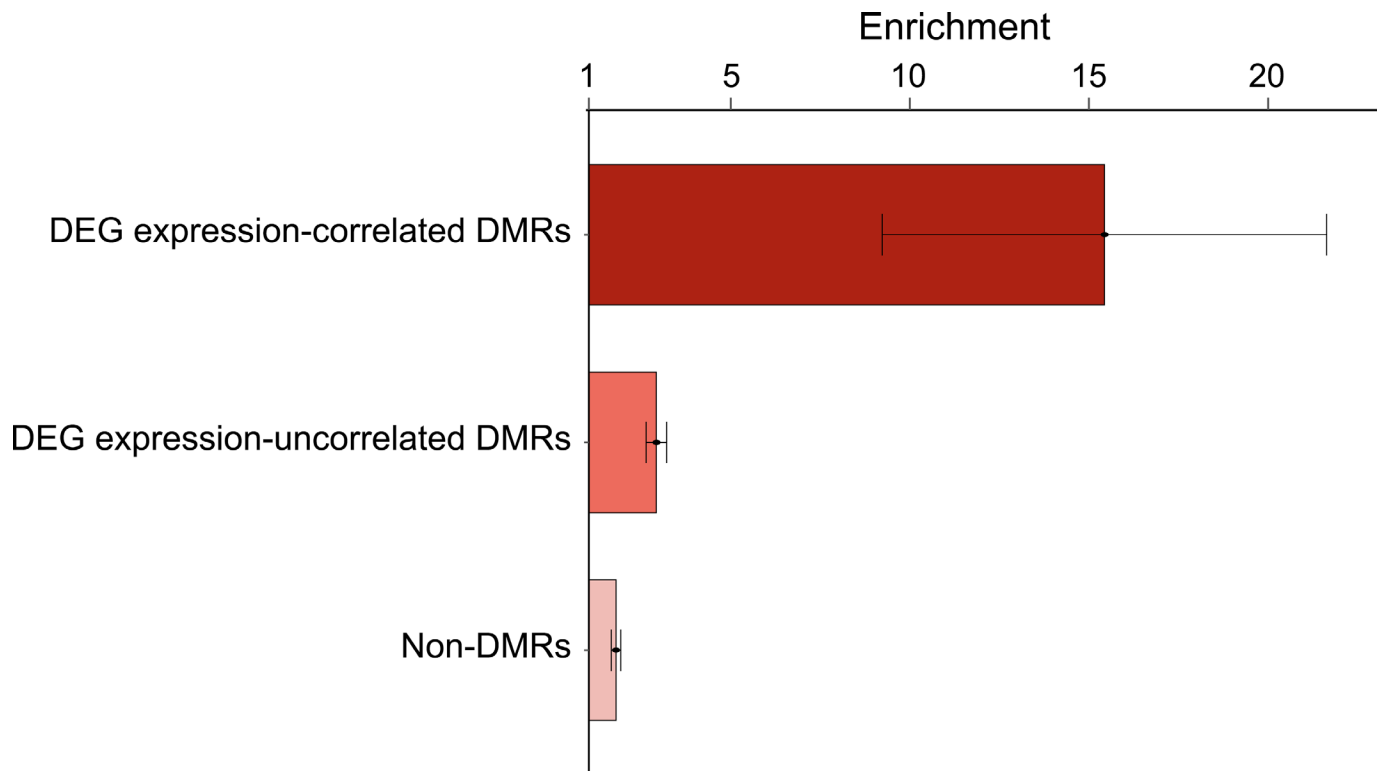
We integrated the individual-level data of DEGs, DMRs and genome-wide genetic variants to understand the regulatory factors underlying the differential expression of DEGs in RA CD4<sup>+</sup> T cells.

A total of 548 RA-specific DEGs (22%) were significantly correlated with RA-specific DMRs in the same TADs or 2 kb windows around their TSSs, at a gene-level FDR of 5%, and eQTLs in this study and other CD4<sup>+</sup> T cell eQTL studies<sup>44 45</sup> were detected in 771 DEGs (figure 4A,B). Half of the DMR-methylation-correlated DEGs showed significantly negative correlations with the DNA methylation levels in DMRs, suggesting bivalency of DNA methylation in transcriptional regulation. Indeed, we found that DMRs were significantly localised in bivalent chromatin states bound to both activating and repressing epigenetic regulators (figure 3E). Among DMR-methylation-correlated DEGs, 83 DEGs were regulated by meQTLs. We statistically confirmed the presence of methylation-mediation effects of meQTLs on 69 DEGs (83.1%; out of the 83 DEGs).

Several known immune genes, involved in diverse immune pathways, were detected as potentially meQTL-regulated DEGs mediated by DNA methylation changes. For example, *CD83* was regulated by rs75943492, mediated by a DMProbe at 100 kb upstream of the TSS (mediation effect = 0.12,  $P_{\text{FDR}} < 5\%$ ; figure 4C).

### Variants in differential expression-associated DMRs that explained significantly more heritability of RA

We observed DEGs in 79 loci out of 118 non-HLA RA-risk loci.<sup>4 5 46</sup> In addition, 43 loci with DEGs had DMRs correlated with expression levels of DEGs in CD4<sup>+</sup> T cells. Considering the highly enriched meQTLs and eQTLs around methylation sites and TSSs, respectively (online supplemental figure S5), we estimated RA heritability explained by all variants in the 5 kb regions around methylation sites or TSSs to examine the potential contribution of DMR-mediated DEG-regulating meQTLs and DEG-regulating eQTLs to RA susceptibility on a genome-wide scale. Strikingly, a relatively large fraction of RA heritability was explained by variants within RA-specific DMRs whose methylation levels correlated with the expression level of DEGs in RA CD4<sup>+</sup> T cells (15.4-fold more heritability than control variants, 8.6-fold more than variants in non-DMRs and 5.3-fold larger than variants in expression-uncorrelated DMRs; figure 5 and online supplemental table S11). The TSS regions with significantly more eQTLs explained a relatively large fraction of RA heritability but the enrichment estimate in DEGs (=33.4) was only 1.6-fold more than that in non-DEGs (=20.5). Although most of the disease association signals are reported to be little correlated with QTL drivers,<sup>33</sup> the genome-wide heritability partitioning analysis strongly suggests a potential regulatory effect of RA variants in DMRs that results in the differential expression of some disease-relevant genes in RA CD4<sup>+</sup> T cells and eventually leads to T cell alteration in patients with RA.



**Figure 5** Large enrichment of partitioned-heritability of RA in DEG expression-correlated DMRs. Enrichment of heritability of RA in three different types of DNA methylated regions was estimated using stratified LDSC. Strong enrichment was detected in DEG expression-correlated DMRs compared with methylated regions irrelevant to DEGs. The SEs of enrichment are shown as error bars. DEGs, differentially expressed genes; DMRs, differential methylated regions; RA, rheumatoid arthritis.

## DISCUSSION

This is the first multi-omics analysis for CD4<sup>+</sup> T cells in patients with RA to profile the plausible causal factors underlying RA-specific DEGs using genomic, transcriptomic and epigenetic data in a single cohort. This study identified the differential expression of several key immune regulator genes involved in the proliferation and differentiation of CD4<sup>+</sup> T cells. A large number of DMRs were identified in genome-wide and targeted methylation quantification approaches, suggesting that DMRs were preferentially located in highly regulatory elements in CD4<sup>+</sup> T cells or TFBSs. A large number of DEGs could be partially explained by meQTL-mediated DMRs or eQTLs and were located in RA susceptibility loci. Nevertheless, it is not common that RA-risk variants are genetically linked in known eQTLs,<sup>33</sup> suggesting a low statistical power in most eQTL analyses and the complex gene regulations involving multiple variants and indirect (epigenetic) regulation. We demonstrated the high enrichment of RA heritability in the region enriched with DMR-mediated DEG-regulating meQTLs.

Our results suggest that several key immune regulators (such as CD83, SMAD7 and GATA3) in CD4<sup>+</sup> T cells are involved in the T-cell alteration in RA. For example, we observed decreased expression of CD83 in patients with RA possibly by a meQTL-mediated DMR. Deficiency of CD83 in mice downregulates the differentiation of T<sub>reg</sub> cells<sup>38</sup> and leads to the proliferation of CD4<sup>+</sup>Foxp3<sup>-</sup> T cells and the differentiation to T<sub>h1</sub> and T<sub>h17</sub> cells, enhancing the immune response.<sup>37</sup> As another example, SMAD7, encoding an inhibitor of TGF-β signalling in T<sub>reg</sub> differentiation,<sup>39</sup> was downregulated in RA CD4<sup>+</sup> T cells. Consistently, a recent study revealed the decreased expression of SMAD7 in synovial tissues of patients with RA as well as severe joint inflammation in a SMAD7-knockout mouse model displaying

imbalanced T<sub>h17</sub>/T<sub>reg</sub> responses.<sup>47</sup> As a third example, GATA3, which is downregulated in RA CD4<sup>+</sup> T cells, is a master regulator of T cell function<sup>48</sup> and a promising target of non-steroidal drugs for the treatment of autoimmune diseases.<sup>49–51</sup>

In conclusion, our findings shed light on how genetic variants can shape the disease-specific transcriptomic signatures in CD4<sup>+</sup> T cells in patients with RA, illustrating the advantage of the same-sample inter-omics data analysis on disease-relevant tissues in dissecting the complex transcriptional regulation driven by genome-wide genetic variants.

## Author affiliations

<sup>1</sup>Department of Biology and Department of Life and Nanopharmaceutical Sciences Biology, Kyung Hee University, Seoul, Republic of Korea

<sup>2</sup>Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, Republic of Korea

<sup>3</sup>Hanyang University Institute for Rheumatology Research, Seoul, Republic of Korea

<sup>4</sup>Division of Genome Science, Department of Precision Medicine, National Institute of Health, Osong Health Technology Administration Complex, Cheongju, Republic of Korea

**Contributors** KK, SCB and BJK designed the study. SYB, HSL and SCB recruited and characterised patients with rheumatoid arthritis and controls. SYB, JHY, JMK, JBB, HSL, BJK, KK and SCB generated genetic, epigenetic and transcriptomic data. EH, JL, KK and SCB analysed the data and interpreted the results. EH, KK and SCB wrote the manuscript. All authors reviewed and approved the manuscript.

**Funding** This study is supported by Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT & Future Planning (2017R1E1A1A01076388), Hanyang University Institute for Rheumatology Research and the National Institute of Health (2012-N73006-01, 2017-NI73002-02). Expression and DNA methylation data were generated as a part of Korean Epigenome Project (4848–308) with the support of the Korea Disease Control and Prevention Agency. KoreanChip was designed by Korean Genome Analysis Project (4845–301) that was supported by the Korea Disease Control and Prevention Agency.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Ethics approval** This study was approved by the Institutional Review Board at Hanyang University (HYG-11-008-1).

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available in a public, open access repository. All data relevant to the study are included in the article or uploaded as supplementary information. The summary statistics of DEGs, DMRs, eQTLs and meQTLs in this study are available at <https://doi.org/10.5061/dryad.w0vt4b8pw>.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

#### ORCID iDs

Eunji Ha <http://orcid.org/0000-0001-8618-8347>

Kwangwoo Kim <http://orcid.org/0000-0001-8926-6216>

Sang-Cheol Bae <http://orcid.org/0000-0003-4658-1093>

#### REFERENCES

- Firestein GS, Budd R, Gabriel SE. *Kelley's Textbook of Rheumatology E-Book*. Elsevier Health Sciences, 2012.
- van der Woude D, Houwing-Duistermaat JJ, Toes REM, et al. Quantitative heritability of anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis. *Arthritis Rheum* 2009;60:916–23.
- Kim K, Bang S-Y, Lee H-S, et al. Update on the genetic architecture of rheumatoid arthritis. *Nat Rev Rheumatol* 2017;13:13–24.
- Kwon Y-C, Lim J, Bang S-Y, et al. Genome-Wide association study in a Korean population identifies six novel susceptibility loci for rheumatoid arthritis. *Ann Rheum Dis* 2020;79:1438–45.
- Ha E, Bae S-C, Kim K. Large-Scale meta-analysis across East Asian and European populations updated genetic architecture and variant-driven biology of rheumatoid arthritis, identifying 11 novel susceptibility loci. *Ann Rheum Dis* 2020;79:558–65.
- Farh KK-H, Marson A, Zhu J, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 2015;518:337–43.
- Finucane HK, Bulik-Sullivan B, Gusev A, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* 2015;47:1228–35.
- Pickrell JK. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am J Hum Genet* 2014;94:559–73.
- Hu X, Kim H, Stahl E, et al. Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. *Am J Hum Genet* 2011;89:496–506.
- Trynka G, Sandor C, Han B, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet* 2013;45:124–30.
- Andersson R, Gebhard C, Miguel-Escalada I, et al. An atlas of active enhancers across human cell types and tissues. *Nature* 2014;507:455–61.
- Ishigaki K, Akiyama M, Kanai M, et al. Large-Scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. *Nat Genet* 2020;52:669–79.
- Moro-Garcia MA, Mayo JC, Sainz RM, et al. Influence of inflammation in the process of T lymphocyte differentiation: proliferative, metabolic, and oxidative changes. *Front Immunol* 2018;9:339.
- Cope AP, Schulze-Koops H, Aringer M. The central role of T cells in rheumatoid arthritis. *Clin Exp Rheumatol* 2007;25:54–11.
- Liu T, Lin X, Yu H. Identifying genes related with rheumatoid arthritis via system biology analysis. *Gene* 2015;571:97–106.
- Chun S, Casparino A, Patsopoulos NA, et al. Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nat Genet* 2017;49:600–5.
- Arnett FC, Edworthy SM, Bloch DA, et al. The American rheumatism association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31:315–24.
- Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43:e47.
- Leek JT, Johnson WE, Parker HS, et al. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012;28:882–3.
- Xiao Y, Hsiao T-H, Suresh U, et al. A novel significance score for gene selection and ranking. *Bioinformatics* 2014;30:801–7.
- Yu G, Wang L-G, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;16:284–7.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–9.
- Zhang Y, Liu T, Meyer CA, et al. Model-Based analysis of ChIP-Seq (MACS). *Genome Biol* 2008;9:R137.
- Stark R, Brown G. *DiffBind: differential binding analysis of ChIP-Seq peak data. R package version.*, 2011: 100, 4–3.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol* 2014;15:550.
- Tian Y, Morris TJ, Webster AP, et al. Champ: updated methylation analysis pipeline for illumina BeadChips. *Bioinformatics* 2017;33:3982–4.
- Teschendorff AE, Marabita F, Lechner M, et al. A beta-mixture quantile normalization method for correcting probe design bias in illumina Infinium 450 K DNA methylation data. *Bioinformatics* 2013;29:189–96.
- Du P, Zhang X, Huang C-C, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 2010;11:587.
- Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* 2017;12:2478–92.
- Kundaje A, Meuleman W, Heravi-Moussavi A. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;518:317–30.
- Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 2015;31:2382–3.
- Sheffield NC, Bock C. Lola: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics* 2016;32:587–9.
- Javierre BM, Burren OS, Wilder SP, et al. Lineage-Specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* 2016;167:e19:1369–84.
- Ongen H, Buil A, Brown AA, et al. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* 2016;32:1479–85.
- Delaneau O, Ongen H, Brown AA, et al. A complete tool set for molecular QTL discovery and analysis. *Nat Commun* 2017;8:15452.
- Tingley D, Yamamoto T, Hirose K, et al. mediation : R Package for Causal Mediation Analysis. *J Stat Softw* 2014;59.
- Liedtke K, Alter C, Günther A, et al. Endogenous CD83 Expression in CD4<sup>+</sup> Conventional T Cells Controls Inflammatory Immune Responses. *J Immunol* 2020;204:3217–26.
- Doebbele M, Koenig C, Krzyzak L, et al. Cd83 expression is essential for Treg cell differentiation and stability. *JCI Insight* 2018;3:jci.insight.99712. doi:10.1172/jci.insight.99712
- Yang L, Pang Y, Moses HL. Tgf-Beta and immune cells: an important regulatory axis in the tumor microenvironment and progression. *Trends Immunol* 2010;31:220–7.
- Giang S, La Cava A. Irf1 and BATF: key drivers of type 1 regulatory T-cell differentiation. *Cell Mol Immunol* 2017;14:652–4.
- Zhao S, Fung-Leung W-P, Bittner A, et al. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One* 2014;9:e78644.
- McRae AF, Marioni RE, Shah S, et al. Identification of 55,000 replicated DNA methylation QTL. *Sci Rep* 2018;8:17605.
- Huffman KM, Jessee R, Andonian B, et al. Molecular alterations in skeletal muscle in rheumatoid arthritis are related to disease activity, physical inactivity, and disability. *Arthritis Res Ther* 2017;19:12.
- Schmiedel BJ, Singh D, Madrigal A, et al. Impact of genetic polymorphisms on human immune cell gene expression. *Cell* 2018;175:e16:1701–15.
- Ishigaki K, Kochi Y, Suzuki A, et al. Polygenic burdens on cell-specific pathways underlie the risk of rheumatoid arthritis. *Nat Genet* 2017;49:1120–5.
- Okada Y, Wu D, Trynka G, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 2014;506:376–81.
- Zhou G, Sun X, Qin Q, et al. Loss of Smad7 promotes inflammation in rheumatoid arthritis. *Front Immunol* 2018;9:2537.
- Wan YY. Gata3: a master of many trades in immune regulation. *Trends Immunol* 2014;35:233–42.
- Barati A, Jamshidi AR, Ahmadi H, et al. Effects of  $\beta$ -d-mannuronic acid, as a novel non-steroidal anti-inflammatory medication within immunosuppressive properties, on *IL17*, *ROR $\gamma$ t*, *IL4* and *GATA3* gene expressions in rheumatoid arthritis patients. *Drug Des Devel Ther* 2017;11:1027–33.
- Nath N, Morinaga O, Singh I. S-Nitrosoglutathione a physiologic nitric oxide carrier attenuates experimental autoimmune encephalomyelitis. *J Neuroimmune Pharmacol* 2010;5:240–51.
- Khadem Azarian S, Jafarnejad-Ansari F, Nazari S, et al. Effects of guluronic acid, as a new NSAID with immunomodulatory properties on *IL-17*, *ROR $\gamma$ t*, *IL-4* and *GATA-3* gene expression in rheumatoid arthritis patients. *Immunopharmacol Immunotoxicol* 2020;42:22–7.