

SCIENTIFIC REPORTS

OPEN

Genetic variants differentially associated with rheumatoid arthritis and systemic lupus erythematosus reveal the disease-specific biology

Jiwoo Lim & Kwangwoo Kim 

Two rheumatic autoimmune diseases, rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE), have distinct clinical features despite their genetic similarities. We hypothesized that disease-specific variants exclusively associated with only one disease could contribute to disease-specific phenotypes. We calculated the strength of disease specificity for each variant in each disease against the other disease using summary association statistics reported in the largest genome-wide association studies of RA and SLE. Most of highly disease-specific associations were explained by non-coding variants that were significantly enriched within regulatory regions (enhancers or H3K4me3 histone modification marks) in specific cell or organ types. (e.g., In RA, regulatory T primary cells, CD4+ memory T primary cells, thymus and lung; In SLE, CD19+ B primary cells, mobilized CD34+ primary cells, regulatory T primary cells and monocytes). Consistently, genes in the disease-specific loci were significantly involved in T cell- and B cell-related gene sets in RA and SLE. In summary, this study identified disease-specific variants between RA and SLE, and provided statistical evidence for disease-specific cell types, organ and gene sets that may drive the disease-specific phenotypes.

Rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE) are well-characterized rheumatic autoimmune diseases with >60% heritability^{1,2}. Genome-wide association studies (GWAS) have uncovered the highly polygenic etiology of RA and SLE, bringing the known disease-associated loci to about 100 in each disease³⁻⁷. The majority of the identified genetic associations are explained by common-frequency variants with modest effect sizes.

A dozen same disease alleles have been detected at the genome-wide significance level in both RA and SLE^{3,4}. For examples, some variants in immune-related genes such as *HLA-DRB1*, *PTPN22*, *STAT4*, and *TNFAIP3* were reported to contribute to risk of RA and SLE⁸⁻¹¹. These pleiotropic variants associated with both RA and SLE strongly suggest that the pathogenesis leading to immune dysfunction is partially shared between these two autoantibody-producing diseases. Evidence of pleiotropic variants in similar diseases motivated cross-disease meta-analyses^{12,13} and phenome-wide association studies to identify pleiotropic variants that explain common pathogenesis in different diseases.

Although RA and SLE share some genetic etiologies, they have highly distinct clinical features in terms of primarily inflamed sites, disease prognosis, and autoantigens. Thus, it is tempting to hypothesize that disease-specific variants exclusively associated with only one disease drive disease-specific features of the disease.

Here, we comprehensively investigated the dissimilarity in disease-variant associations between RA and SLE at the genome-wide level, and identified highly disease-specific variants that map to disease-specific cell types and pathways.

Department of Biology, Kyung Hee University, Seoul, Republic of Korea. Correspondence and requests for materials should be addressed to K.K. (email: kkim@khu.ac.kr)

Methods

GWAS summary association statistic data. The largest-ever European GWAS summary association statistic data for RA and SLE were obtained from Okada *et al.*³ and Bentham *et al.*⁴, respectively. Both datasets were generated after whole-genome imputation using the reference panel from the 1000 Genomes Project data¹⁴ followed by disease association tests using multivariable logistic regression adjusting for population stratifications. The summary statistics for disease association include odds ratios of effect alleles, their standard errors (or 95% confidence intervals), effect alleles, non-effect alleles, association *P* values, and imputation quality in autosomal single nucleotide polymorphisms (SNPs). All subsequent analyses were performed for the SNPs with minor allele frequency (MAF) > 0.5% and imputed score > 0.5 in RA and/or SLE datasets.

Meta-analysis. A cross-disease meta-analysis was performed using GWAMA software¹⁵, which calculate the inverse of variances to weight effect sizes, based on the fixed-effects model. The heterogeneity of association effect estimates between RA and SLE was assessed using Cochran's Q test.

Estimating the strength of disease specificity. We created a statistic (*S*) that indicates the strength of disease-specific association of a variant in a disease using the following equation:

$$\text{Strength of disease specificity } (S_{ij}) = \left| \frac{\beta_{ij}}{se_{ij}} \right| \times Z_i,$$

where S_{ij} is the disease specificity of SNP *i* in disease *j*; β_{ij} is the effect size of the effect allele in the SNP *i* in disease *j*; se_{ij} is the standard error of β_{ij} ; and Z_i is the z-score transformed from *P* values in Cochran's Q test which indicates the effect-size heterogeneity of SNP *i* between RA and SLE. The S_{ij} statistic is calculated by multiplying the absolute value of a *t*-statistic (equal to β_{ij} divided by se_{ij}) by Z_i . Thus, the *S* statistic is synergistically increased only when the association of a tested SNP is strong in a disease and the effect estimates between two diseases are highly heterogeneous. The near-zero value of the *t*-statistic and/or *Z*-score results in weak or mild *S* estimates. Each variant generates two *S* statistics (each for each disease); we compared the two *S* statistics from each variant and used the larger *S* value for the subsequent analyses. Then, all variants were ranked by the strength of disease specificity for each disease. Finally, the top 1% of SNPs in each disease were extracted to use in all subsequent enrichment analyses for each disease.

Enrichment analysis using H3K4me3 histone modification marks. An enrichment analysis was performed using the Epi-GWAS software^{14,16} to determine whether disease-specific variants significantly overlap with histone posttranslational modification marks in specific cell types. Among the various types of histone marks, H3K4me3 is well-known as the most cell-type specific histone modification mark¹⁶. Briefly, the enrichment score was calculated based on the positions of query and proxy SNPs ($r^2 > 0.8$) within H3K4me3 peaks for each cell type (obtained from chromatin immunoprecipitation sequencing in the Roadmap Epigenomics Project)^{17,18}, taking into account the height and position of peak summits. Then, a permutation test was performed to identify the cell types that exhibited significantly high overlap between the H3K4me3 marks and disease-specific variants.

Enrichment analysis using enhancers. An enrichment analysis was also performed using enhancer annotation data for 71 cell types and 41 organ types obtained from the FANTOM5 consortium database^{19,20}. An enhancer-by-cell type matrix or enhancer-by-organ matrix filled with 0 (=no enhancer) or 1 (=enhancer) was constructed, and the matrix values were randomly shuffled by cell or organ type 10,000 times, preserving the total number of the enhancers in each cell type or organ. The expected distribution of the numbers of enhancers and disease-specific SNPs in each cell or organ type was obtained by counting the shuffled enhancers in which the query and proxy ($r^2 > 0.8$) SNPs were located. Thus, this procedure did not disrupt the important property such as the total number of enhancers in each cell type, the number of appearances of each enhancer across different cell types, and the linkage disequilibrium among the query SNPs. The significance value of overlaps in each of cell or organ types was calculated by determining the percentile rank of the actual number of the overlap between enhancers and disease-specific SNPs in the tested cell or organ types among the numbers of the overlaps observed from the 10,000 shuffled matrices.

Gene set enrichment analysis. A gene set enrichment analysis was conducted using DEPICT software^{14,21} to identify biological pathways implicated by genes in disease-specific loci that were exclusively associated with only one disease. Briefly, the top 1% of disease-specific SNPs (pruned by $r^2 < 0.2$ at *P* in a cross-disease meta-analysis < 0.05) and their percentile ranking in each disease were used as query variants and *P* values, respectively. The DEPICT software uses the customized gene sets that were previously generated from existing gene set databases^{22–26} and the co-expression data from human microarrays²⁷. Genes that were mapped with neighboring disease-specific SNPs in each disease were tested for enrichments in the customized biological pathways based on a false-discovery rate (FDR) threshold of 20%.

Results

Numerous disease-risk variants shared between RA and SLE. To explore the disease-specific associations between RA and SLE, we obtained the association summary statistics from the largest-ever GWAS in European populations. The study populations consisted of 43,923 controls and 14,361 cases in the RA GWAS, and 6,959 controls and 4,036 cases in the SLE GWAS. A total of 8,031,027 autosomal SNPs with MAF \geq 0.5% and imputation quality score \geq 0.5 in both RA and SLE datasets were analyzed in this study.

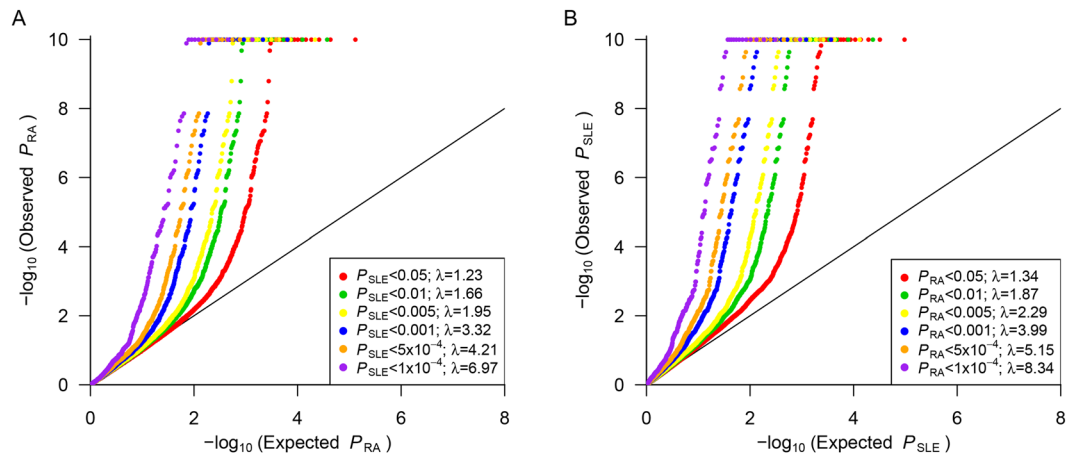


Figure 1. (A) Quantile-quantile plot showing inflated RA associations of SLE-associated SNPs with $P_{SLE} < 0.05$, 0.01, 0.005, 0.001, 5×10^{-4} or 1×10^{-4} . (B) Quantile-quantile plot showing inflated SLE associations of RA-associated SNPs with $P_{RA} < 0.05$, 0.01, 0.005, 0.001, 5×10^{-4} or 1×10^{-4} . The diagonal red line represents perfect concordance of observed and expected $-\log_{10}P$ values.

First, before focusing on association differences in RA and SLE, we evaluated a degree of risk-allele sharing between two diseases. To avoid inflated estimates of the association similarity or dissimilarity in two diseases due to multiple SNPs in linkage disequilibrium (LD), we extracted 369,955 SNPs that were not correlated with each other ($r^2 < 0.2$) within a 2-Mb flanking window based on the genotypes in the European populations of the 1000 Genomes Projects. The SNP-disease association P values displayed a significantly positive correlation between RA and SLE in the logarithm scale with a plus or minus sign representing the positive or negative effect of a minor allele, respectively ($P < 2.2 \times 10^{-16}$; Supplementary Fig. S1). These results suggest that both diseases share a number of disease-risk variants.

We identified highly inflated associations of SNPs for one disease (either RA or SLE) that were associated with the other disease (either SLE or RA) at various significance threshold levels, as shown in conditional quantile-quantile plots²⁸ in Fig. 1. For example, variants with an RA-association P value < 0.05 showed inflated SLE associations compared to those expected based on randomly selected SNPs (inflation factor $\lambda = 1.34$). When the same analysis was performed using variants with more significant P values in RA (e.g., $P < 10^{-4}$), we found more inflated associations with SLE ($\lambda = 8.34$). Similarly, SLE variants passing various association P value thresholds were extracted and evaluated for their associations with RA. The results clearly indicated that the association with RA became highly inflated as the P value thresholds for SLE associations declined (Fig. 1). In addition, a pleiotropy analysis for all pruned SNPs with $r^2 < 0.2$ in RA and SLE using GPA^{29,30} consistently supported a high degree of disease-allele sharing ($P < 1.0 \times 10^{-100}$).

We identified 14 non-HLA loci surpassing the genome-wide significance level in a cross-disease meta-analysis ($2.43 \times 10^{-12} \leq P_{\text{meta}} \leq 4.48 \times 10^{-8}$) but not in a single disease association analysis (Supplementary Table S1). Of them, 13 regional associations were validated in additional samples in the same studies^{3,4} or recent independent studies^{5,6,31,32}, which also indicates the RA and SLE have highly similar genetic architectures. (One locus with SNP rs8045689 that have not yet been significantly associated with RA or SLE could be a good candidate of a pleiotropic disease-associated variant).

Identification of disease-specific loci. We extracted 64,845 SNPs from both the RA and SLE datasets that were not correlated with each other ($r^2 < 0.2$). We calculated the strength of disease specificity (S) for each SNP using the disease-specific effect estimate, its standard error, and heterogeneity of effect estimates between two diseases (details in METHODS). Highly RA- and SLE-specific SNPs are listed in Supplementary Tables S2 and S3, respectively. For example, the SNP rs34185821 is located 80 kb from *RBPJ* and displayed one of the highest scores of disease specificity in RA ($P_{RA} = 2.20 \times 10^{-16}$ and $OR_{RA} = 0.87$ in RA; $P_{SLE} = 0.419$ and $OR_{SLE} = 1.02$ in SLE), whereas the intergenic SNP rs13019891 was highly SLE-specific ($P_{RA} = 0.90$ and $OR_{RA} = 1.00$ in RA; $P_{SLE} = 2.26 \times 10^{-36}$ and $OR_{SLE} = 1.86$ in SLE).

We selected the top 1% of the most disease-specific non-major histocompatibility complex (non-MHC) SNPs in RA and SLE, and then retrieved their proximal SNPs that were correlated with the disease-specific SNPs ($r^2 > 0.8$ in European populations in the 1000 Genomes Project). Finally, a total of 4,913 RA-specific SNPs and 8,223 SLE-specific SNPs were used for all subsequent analyses. Most of these variants were found in non-coding regions, and 3.5% of loci (23 loci out of 648) had at least one leading or proxy variant that altered amino acid residues. Therefore, we postulated that the disease-specific SNPs drive the disease-specific phenotypes primarily by regulating gene expression rather than regulating protein activity.

Specific cell or organ types implicated by disease-specific variants. As the majority of the disease-specific loci were explained by non-coding variants, we hypothesized that disease-specific allele affected the regulatory elements (e.g. histone marks and enhancers), which are highly cell type-specific. Under this assumption, we aimed to identify disease-relevant cell or organ types whose regulatory regions should be

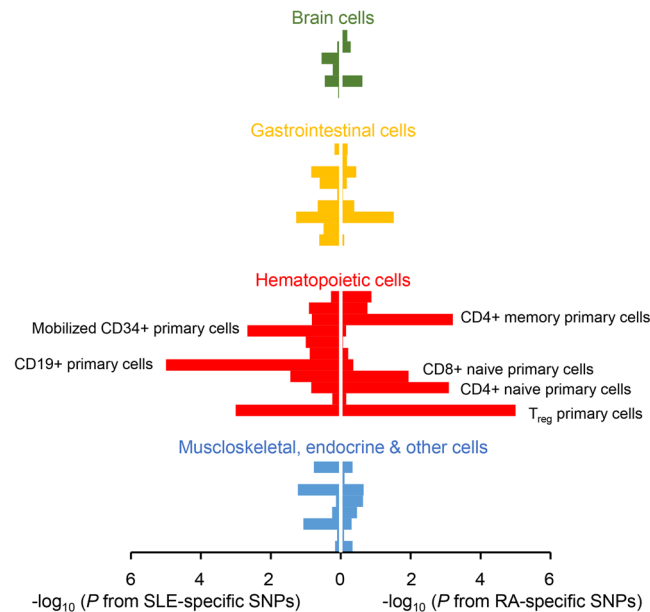


Figure 2. Significant overlaps between cell type-specific H3K4me3 histone modification marks and disease-specific variants. Enrichment P values of RA- and SLE-specific variants within H3K4me3 marks were plotted for each cell type in a minus \log_{10} scale. The x-axis represents the transformed enrichment P values for SLE variants (left) and RA variants (right). The bars for each cell type were colored according to the lineage classification of the cell type. The names of tissues with enrichment P values < 0.05 were labeled next to the corresponding bar. The detailed information of this figure is shown in Supplementary Table S4.

enriched with the disease-specific variants. We annotated each disease-specific variant and its proxy variants ($r^2 > 0.8$) for the following potential regulatory elements: (1) the most cell type-specific histone modification marks (H3K4me3) in 34 cell types from the NIH Roadmap epigenomics consortium^{17,18} and (2) enhancer regions in 71 cell types and 41 organ types from the FANTOM5 consortium^{19,20}.

The results showed that H3K4me3 peaks in regulatory T primary cells were significantly colocalized with RA- and SLE-specific SNPs ($P < 1.0 \times 10^{-5}$ and $P = 9.9 \times 10^{-4}$, respectively), although disease-specific SNPs were mutually exclusive between two diseases. This suggests that regulatory T cells may contribute to the RA and SLE pathogenesis via different genes and/or pathways, or via alternative expressional regulation of the same genes. In addition to regulatory T cells, other T-cell subtypes including CD4⁺ memory T primary cells and CD4⁺ naive T primary cells were implicated in RA by significant H3K4me3 localization on RA-specific SNPs ($P \leq 8.1 \times 10^{-4}$). In SLE, CD19⁺ B primary cells and mobilized CD34⁺ primary cells showed significant enrichments ($P < 1 \times 10^{-5}$ and $P = 2.14 \times 10^{-3}$, respectively; Fig. 2, Supplementary Table S4).

The disease-specific SNPs were evaluated for the enrichment in the cell-specific enhancer regions in various cell and organ types. We found significant overlap between RA-specific SNPs and enhancers in three immune cell types ($P = 0.010$ in dendritic cells; $P = 0.011$ in NK cells; $P = 0.027$ in T cells) and two organs ($P = 0.0052$ in thymus; $P = 0.015$ in lung; Fig. 3, Supplementary Table S5). These analyses consistently support the importance of T-cell biology in RA³³ and the disease relevance of lung in RA. Lung is considered as a key organ in the initiation of RA-specific immune responses, especially for the production of anti-citrullinated protein antibodies (ACPA)³⁴.

SLE-specific SNPs were significantly more in the enhancer regions in monocyte ($P = 0.010$), granulocyte ($P = 0.016$) and basophil ($P = 0.018$). We note the association of B cells was marginal ($P = 0.059$) in SLE. In an organ-enhancer analysis, the blood enhancer regions were significantly colocalized with the SLE-specific SNPs ($P < 1.0 \times 10^{-4}$; Fig. 3, Supplementary Table S5). The association of meninx related to psychiatric SLE approached the borderline of significance in this analysis ($P = 0.068$)³⁵.

Specific gene sets implicated by disease-specific alleles. It is plausible that genes in disease-specific loci are involved in disease-relevant biological pathways³⁶. Therefore, we performed a gene-set enrichment analysis for genes near the top 1% of disease-specific SNPs using the DEPICT software²¹ and its own reconstituted gene sets. Many immune-related gene sets and some non-immune gene sets contained significantly more genes in disease-specific loci, with 236 pathways in RA and 199 pathways in SLE at false discovery rate (FDR) $< 20\%$ (Supplementary Tables S6 and S7).

T-cell-related gene sets were most highlighted in RA, including ‘abnormal CD8⁺ T-cell physiology’ ($P = 9.68 \times 10^{-6}$), ‘abnormal lymphocyte morphology’ ($P = 3.06 \times 10^{-5}$), and ‘increased T-cell number’ ($P = 3.20 \times 10^{-5}$). By contrast, other lymphocyte-related gene sets were most emphasized in SLE, including ‘glucose-6-phosphate isomerase protein-protein interaction (GPI PPI subnetwork, $P = 2.08 \times 10^{-5}$)’, ‘TRAF3 PPI subnetwork’ ($P = 3.82 \times 10^{-5}$) and ‘decreased interleukin-2 secretion’ ($P = 4.57 \times 10^{-5}$). The GPI PPI subnetwork is known as neuroleukin or lymphokine that induces antibody secretion of B-cells³⁷. In addition, TRAF3 PPI subnetwork

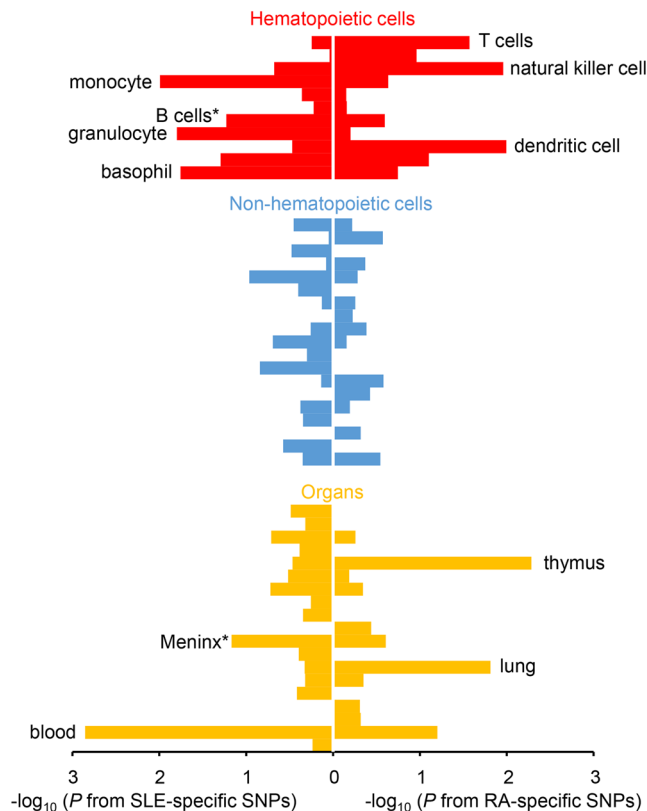


Figure 3. Significant overlaps between cell or organ type-specific enhancers and disease-specific variants. Enrichment P values of RA- and SLE-specific variants within enhancer regions were plotted for each cell or organ type in a minus \log_{10} scale. The x-axis represents the transformed enrichment P values for SLE variants (left) and RA variants (right). Each cell or bar type were color categorized as hematopoietic cell, non-hematopoietic cell, or organ. The names of tissues or organs with enrichment P values < 0.05 were labeled next to the corresponding bar. Tissues or organs with $P > 0.6$ for both RA- and SLE-specific variants have been omitted in this figure. The detailed information of this figure is shown in Supplementary Table S5.

is involved in regulating B cell signaling and plasma cell development³⁸. Many immune-related gene sets were shared in both RA and SLE, although their significance ranks varied (Supplementary Tables S6 and S7).

Discussion

Our genome-wide approach showed that RA and SLE share highly similar genetic etiologies with some exceptional variants that may contribute to disease risk in a disease-specific manner. To evaluate the disease specificity of genetic associations between RA and SLE, we calculated the disease-specificity statistic S for each variant using disease association summary statistics in each disease and association heterogeneity statistics between two diseases. (We note that the statistic S is developed for a prioritization analysis but not for a parametric analysis. The statistic S dependent on MAFs of variants and sample sizes that largely affect standard errors of the disease effect size. Therefore, a null distribution of S is study-specific).

In our analyses, most of the disease-specific associations were explained by non-coding variants, suggesting that disease-specific variants have regulatory effects on regulatory annotations. As the regulatory regions such as histone modification and enhancers are highly cell and organ type-specific, we could trace which cell types or tissues are relevant to each disease by analyzing the overlap between regulatory regions and disease-specific variants. Additionally, genes within or near the disease-specifically associated loci could suggest which gene sets or biological pathways were associated with disease-specific outcomes.

In a series of the enrichment analyses using the top 1% of disease-specific variants in each disease, the importance of T cells in RA pathogenesis and B cells in SLE pathogenesis was identified by H3K4me3. This finding is supported by previous genetic and immunological studies^{39–44}. Notably, regulatory T primary cells were identified in both diseases, although disease-specific variants did not overlap. This implicates that regulatory T cells may have a role in the pathogenesis of RA and SLE with their disease-specific pathways.

A permutation analysis for cell and organ type-specific enhancers, we detected additional cell types and tissues involved in RA and SLE pathogenesis. The significance values were relatively weak due to small coverage of the entire enhancer regions in the human genome that contained relatively few disease-specific SNPs. Enhancers with RA-specific variants supported a role for the antibody-mediated immune systems⁴⁵, whereas enhancers with SLE-specific variants were significantly overlapped with enhancer regions in the monocytes that contribute to innate immunity and subsequent antigen presentation⁴⁶. SLE-specific variants also were associated with

granulocytes and basophils, which interact with B cells in SLE pathogenesis. Interestingly, we found the association of lung in RA, where ACPA may be produced in patients with RA^{47,48}.

Gene set enrichment analyses using disease-specific variants indicated that the majority of pathways detected at FDR < 20% were related to immune response and cytokines. Most their effects were considered as valid in the cell types that were identified in the enrichment analysis for histone marks and enhancers (e.g., T cell-related gene sets in RA; B cell-related and immunoglobulin-related gene sets in SLE; lung inflammation in RA). These results were consistent with the enrichment results for cell- and organ specific enhancers.

In conclusions, this work illustrated the advantage of identifying disease-specific variants between two similar diseases in understanding disease-specific cell types, organ and biological pathways. We identified disease-specific variants between two rheumatic diseases, RA and SLE, and provided statistical evidence for disease-specific cells, tissues, and gene sets that may drive the distinctly different disease-specific phenotypes.

References

- Moser, K. L., Kelly, J. A., Lessard, C. J. & Harley, J. B. Recent insights into the genetic basis of systemic lupus erythematosus. *Genes Immun* **10**, 373–379, <https://doi.org/10.1038/gene.2009.39> (2009).
- Smolen, J. S., Aletaha, D. & McInnes, I. B. Rheumatoid arthritis. *Lancet* **388**, 2023–2038, [https://doi.org/10.1016/S0140-6736\(16\)30173-8](https://doi.org/10.1016/S0140-6736(16)30173-8) (2016).
- Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381, <https://doi.org/10.1038/nature12873> (2014).
- Bentham, J. *et al.* Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat Genet* **47**, 1457–1464, <https://doi.org/10.1038/ng.3434> (2015).
- Morris, D. L. *et al.* Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus. *Nat Genet* **48**, 940–946, <https://doi.org/10.1038/ng.3603> (2016).
- Stahl, E. A. *et al.* Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet* **42**, 508–514, <https://doi.org/10.1038/ng.582> (2010).
- Eyre, S. *et al.* High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat Genet* **44**, 1336–1340, <https://doi.org/10.1038/ng.2462> (2012).
- Stanford, S. M. & Bottini, N. PTPN22: the archetypal non-HLA autoimmunity gene. *Nat Rev Rheumatol* **10**, 602–611, <https://doi.org/10.1038/nrrheum.2014.109> (2014).
- Shimane, K. *et al.* The association of a nonsynonymous single-nucleotide polymorphism in TNFAIP3 with systemic lupus erythematosus and rheumatoid arthritis in the Japanese population. *Arthritis Rheum* **62**, 574–579, <https://doi.org/10.1002/art.27190> (2010).
- Ji, J. D. *et al.* Association of STAT4 polymorphism with rheumatoid arthritis and systemic lupus erythematosus: a meta-analysis. *Mol Biol Rep* **37**, 141–147, <https://doi.org/10.1007/s11033-009-9553-z> (2010).
- Shimane, K. *et al.* An association analysis of HLA-DRB1 with systemic lupus erythematosus and rheumatoid arthritis in a Japanese population: effects of *09:01 allele on disease phenotypes. *Rheumatology (Oxford)* **52**, 1172–1182, <https://doi.org/10.1093/rheumatology/kes427> (2013).
- Marquez, A. *et al.* A combined large-scale meta-analysis identifies COG6 as a novel shared risk locus for rheumatoid arthritis and systemic lupus erythematosus. *Ann Rheum Dis* **76**, 286–294, <https://doi.org/10.1136/annrheumdis-2016-209436> (2017).
- Orozco, G. *et al.* Study of the common genetic background for rheumatoid arthritis and systemic lupus erythematosus. *Ann Rheum Dis* **70**, 463–468, <https://doi.org/10.1136/ard.2010.137174> (2011).
- Genomes Project, C. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073, <https://doi.org/10.1038/nature09534> (2010).
- Magi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11**, 288, <https://doi.org/10.1186/1471-2105-11-288> (2010).
- Trynka, G. *et al.* Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet* **45**, 124–130, <https://doi.org/10.1038/ng.2504> (2013).
- Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330, <https://doi.org/10.1038/nature14248> (2015).
- Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* **28**, 1045–1048, <https://doi.org/10.1038/nbt1010-1045> (2010).
- Lizio, M. *et al.* Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol* **16**, 22, <https://doi.org/10.1186/s13059-014-0560-6> (2015).
- Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461, <https://doi.org/10.1038/nature12787> (2014).
- Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890, <https://doi.org/10.1038/ncomms6890> (2015).
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* **40**, D109–114, <https://doi.org/10.1093/nar/gkr988> (2012).
- Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25–29, <https://doi.org/10.1038/75556> (2000).
- Croft, D. *et al.* Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res* **39**, D691–697, <https://doi.org/10.1093/nar/gkq1018> (2011).
- Lage, K. *et al.* A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol* **25**, 309–316, <https://doi.org/10.1038/nbt1295> (2007).
- Blake, J. A. *et al.* The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse. *Nucleic Acids Res* **42**, D810–817, <https://doi.org/10.1093/nar/gkt1225> (2014).
- Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res* **41**, D991–995, <https://doi.org/10.1093/nar/gks1193> (2013).
- Andreassen, O. A. *et al.* Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. *PLoS Genet* **9**, e1003455, <https://doi.org/10.1371/journal.pgen.1003455> (2013).
- Chung, D., Yang, C., Li, C., Gelernter, J. & Zhao, H. GPA: a statistical approach to prioritizing GWAS results by integrating pleiotropy and annotation. *PLoS Genet* **10**, e1004787, <https://doi.org/10.1371/journal.pgen.1004787> (2014).
- Wang, Q., Yang, C., Gelernter, J. & Zhao, H. Pervasive pleiotropy between psychiatric disorders and immune disorders revealed by integrative analysis of multiple GWAS. *Hum Genet* **134**, 1195–1209, <https://doi.org/10.1007/s00439-015-1596-8> (2015).
- Walsh, A. M. *et al.* Integrative genomic deconvolution of rheumatoid arthritis GWAS loci into gene and cell type associations. *Genome Biol* **17**, 79, <https://doi.org/10.1186/s13059-016-0948-6> (2016).
- Jin, Y. *et al.* Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo. *Nat Genet* **44**, 676–680, <https://doi.org/10.1038/ng.2272> (2012).

33. Cope, A. P. T cells in rheumatoid arthritis. *Arthritis Res Ther* **10**(Suppl 1), S1, <https://doi.org/10.1186/ar2412> (2008).
34. McInnes, I. B. & Schett, G. The pathogenesis of rheumatoid arthritis. *N Engl J Med* **365**, 2205–2219, <https://doi.org/10.1056/NEJMra1004965> (2011).
35. Stock, A. D., Gelb, S., Pasternak, O., Ben-Zvi, A. & Putterman, C. The blood brain barrier and neuropsychiatric lupus: new perspectives in light of advances in understanding the neuroimmune interface. *Autoimmun Rev* **16**, 612–619, <https://doi.org/10.1016/j.autrev.2017.04.008> (2017).
36. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**, 15545–15550, <https://doi.org/10.1073/pnas.0506580102> (2005).
37. Gurney, M. E. *et al.* Neuroleukin: a lymphokine product of lectin-stimulated T cells. *Science* **234**, 574–581 (1986).
38. Lin, W. W. *et al.* The adaptor protein TRAF3 inhibits interleukin-6 receptor signaling in B cells to limit plasma cell development. *Sci Signal* **8**, ra88, <https://doi.org/10.1126/scisignal.aaa5157> (2015).
39. Zhang, P. & Lu, Q. Genetic and epigenetic influences on the loss of tolerance in autoimmunity. *Cell Mol Immunol*, <https://doi.org/10.1038/cmi.2017.137> (2018).
40. Morel, L. Immunometabolism in systemic lupus erythematosus. *Nat Rev Rheumatol* **13**, 280–290, <https://doi.org/10.1038/nrrheum.2017.43> (2017).
41. Ishigaki, K. *et al.* Polygenic burdens on cell-specific pathways underlie the risk of rheumatoid arthritis. *Nat Genet* **49**, 1120–1125, <https://doi.org/10.1038/ng.3885> (2017).
42. Raj, T. *et al.* Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* **344**, 519–523, <https://doi.org/10.1126/science.1249547> (2014).
43. Kim, K., Bang, S. Y., Lee, H. S. & Bae, S. C. Update on the genetic architecture of rheumatoid arthritis. *Nat Rev Rheumatol* **13**, 13–24, <https://doi.org/10.1038/nrrheum.2016.176> (2017).
44. Jarvinen, T. M. *et al.* Replication of GWAS-identified systemic lupus erythematosus susceptibility genes affirms B-cell receptor pathway signalling and strengthens the role of IRF5 in disease susceptibility in a Northern European population. *Rheumatology (Oxford)* **51**, 87–92, <https://doi.org/10.1093/rheumatology/ker263> (2012).
45. Freudenberg, J., Gregersen, P. & Li, W. Enrichment of Genetic Variants for Rheumatoid Arthritis within T-Cell and NK-Cell Enhancer Regions. *Mol Med* **21**, 180–184, <https://doi.org/10.2119/molmed.2014.00252> (2015).
46. Shi, L. *et al.* Monocyte enhancers are highly altered in systemic lupus erythematosus. *Epigenomics* **7**, 921–935, <https://doi.org/10.2217/epi.15.47> (2015).
47. Perry, E., Kelly, C., Eggleton, P., De Soyza, A. & Hutchinson, D. The lung in ACPA-positive rheumatoid arthritis: an initiating site of injury? *Rheumatology (Oxford)* **53**, 1940–1950, <https://doi.org/10.1093/rheumatology/keu195> (2014).
48. Sparks, J. A. & Karlson, E. W. The Roles of Cigarette Smoking and the Lung in the Transitions Between Phases of Preclinical Rheumatoid Arthritis. *Curr Rheumatol Rep* **18**, 15, <https://doi.org/10.1007/s11926-016-0563-2> (2016).

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT & Future Planning (2017R1E1A1A01076388) and Korea Healthcare Technology R & D Project funded by the Ministry for Health & Welfare (HI15C3182) in Republic of Korea.

Author Contributions

J.L. and K.K. contributed to the design and implementation of the research and wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-39132-2>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019