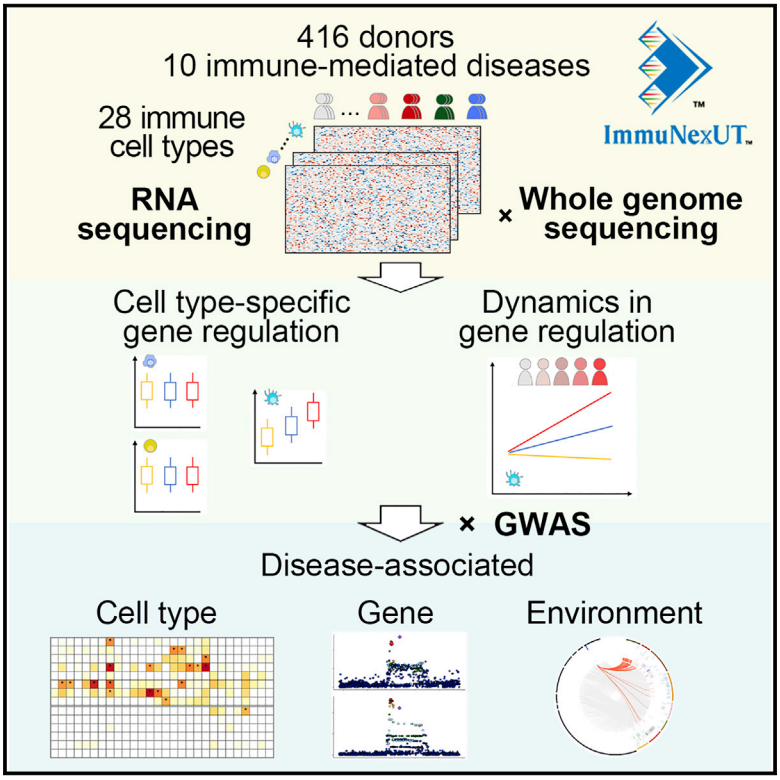
Cell

Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases

Graphical abstract



Highlights

- Gene-regulation atlas of 28 immune cell types under immune-mediated diseases (IMDs)
- Expression QTLs show immune cell-type and disease context specificity
- Cellular pathways diversify eQTL effects under *in vivo* disease conditions
- This atlas links IMD GWAS variants to susceptible genes, cell types, and environment

Authors

Mineto Ota, Yasuo Nagafuchi, Hiroaki Hatano, ..., Kazuhiko Yamamoto, Tomohisa Okamura, Keishi Fujio

Correspondence

otam-int@h.u-tokyo.ac.jp (M.O.), fujiok-int@h.u-tokyo.ac.jp (K.F.)

In brief

The gene-regulation atlas of 28 immune cell types was constructed with immune-mediated disease patient samples and shows the dynamics of gene regulation depending on cell types and immunological conditions.

Ota et al., 2021, Cell *184*, 1–16
May 27, 2021
https://doi.org/10.1016/j.cell.2021.03.056

CellPress

Please cite this article in press as: Ota et al., Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases, Cell (2021), https://doi.org/10.1016/j.cell.2021.03.056

Cell

CellPress

Resource

Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases

Mineto Ota,[1,2,*] Yasuo Nagafuchi,[1,2,14] Hiroaki Hatano,[1,14] Kazuyoshi Ishigaki,[3] Chikashi Terao,[4] Yusuke Takeshima,[1,2] Haruyuki Yanaoka,[1] Satomi Kobayashi,[1] Mai Okubo,[1] Harumi Shirai,[1] Yusuke Sugimori,[1] Junko Maeda,[1] Masahiro Nakano,[1] Saeko Yamada,[1] Ryochi Yoshida,[1] Haruka Tsuchiya,[1] Yumi Tsuchida,[1] Shuji Akizuki,[5] Hajime Yoshifuji,[5] Koichiro Ohmura,[5] Tsuneyo Mimori,[5] Ken Yoshida,[6] Daitaro Kurosaka,[6] Masato Okada,[7] Keigo Setoguchi,[8] Hiroshi Kaneko,[9] Nobuhiro Ban,[10] Nami Yabuki,[10] Kosuke Matsuki,[10] Hironori Mutoh,[10] Sohei Oyama,[10] Makoto Okazaki,[10] Hiroyuki Tsunoda,[10] Yukiko Iwasaki,[1] Shuji Sumitomo,[1] Hirofumi Shoda,[1] Yuta Kochi,[11,12] Yukinori Okada,[13] Kazuhiko Yamamoto,[1,12] Tomohisa Okamura,[2] and Keishi Fujio[1,15,*]

[1]Department of Allergy and Rheumatology, Graduate School of Medicine, The University of Tokyo, Tokyo 113-0033, Japan
[2]Department of Functional Genomics and Immunological Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo 113-0033, Japan
[3]Center for Data Sciences, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA
[4]Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa 230-0045, Japan
[5]Department of Rheumatology and Clinical Immunology, Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan
[6]Division of Rheumatology, Department of Internal Medicine, The Jikei University School of Medicine, Tokyo 105-8461, Japan
[7]Immuno-Rheumatology Center, St. Luke's International Hospital, Tokyo 104-8560, Japan
[8]Division of Collagen Disease, Department of Medicine, Tokyo Metropolitan Komagome Hospital, Tokyo 113-0021, Japan
[9]Division of Rheumatic Diseases, National Center for Global Health and Medicine, Tokyo 162-8655, Japan
[10]Research Division, Chugai Pharmaceutical Co., Ltd., Kamakura, Kanagawa 247-8530, Japan
[11]Department of Genomic Function and Diversity, Medical Research Institute, Tokyo Medical and Dental University, Tokyo 113-8510, Japan
[12]Laboratory for Autoimmune Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa 230-0045, Japan
[13]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Osaka 565-0871, Japan
[14]These authors contributed equally
[15]Lead contact
*Correspondence: otam-int@h.u-tokyo.ac.jp (M.O.), fujiok-int@h.u-tokyo.ac.jp (K.F.)
https://doi.org/10.1016/j.cell.2021.03.056

SUMMARY

Genetic studies have revealed many variant loci that are associated with immune-mediated diseases. To elucidate the disease pathogenesis, it is essential to understand the function of these variants, especially under disease-associated conditions. Here, we performed a large-scale immune cell gene-expression analysis, together with whole-genome sequence analysis. Our dataset consists of 28 distinct immune cell subsets from 337 patients diagnosed with 10 categories of immune-mediated diseases and 79 healthy volunteers. Our dataset captured distinctive gene-expression profiles across immune cell types and diseases. Expression quantitative trait loci (eQTL) analysis revealed dynamic variations of eQTL effects in the context of immunological conditions, as well as cell types. These cell-type-specific and context-dependent eQTLs showed significant enrichment in immune disease-associated genetic variants, and they implicated the disease-relevant cell types, genes, and environment. This atlas deepens our understanding of the immunogenetic functions of disease-associated variants under *in vivo* disease conditions.

INTRODUCTION

Immune-mediated diseases (IMDs) consist of a wide range of pathologies from autoimmunity to autoinflammation (van Kempen et al., 2015), and they affect a large number of people around the world. Although aberrant inflammatory cytokine activity in IMDs indicates the altered functions of immune cells (Wahren-Herlenius and Dörner, 2013), little is known about the responsible genes or even relevant cell types underlying IMDs.

Genome-wide association studies (GWAS) have revealed a number of genomic loci that are associated with complex traits and diseases, including IMDs (Liu et al., 2015; Okada et al., 2014). The majority of these variants lie in non-coding genomic regions (Farh et al., 2015), especially regulatory elements such as enhancers and promoters (Maurano et al., 2012). To decipher the function of these regulatory variants, we sought to examine their effects on gene regulation by analysis of expression quantitative trait loci (eQTL).

Please cite this article in press as: Ota et al., Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases, Cell (2021), https://doi.org/10.1016/j.cell.2021.03.056

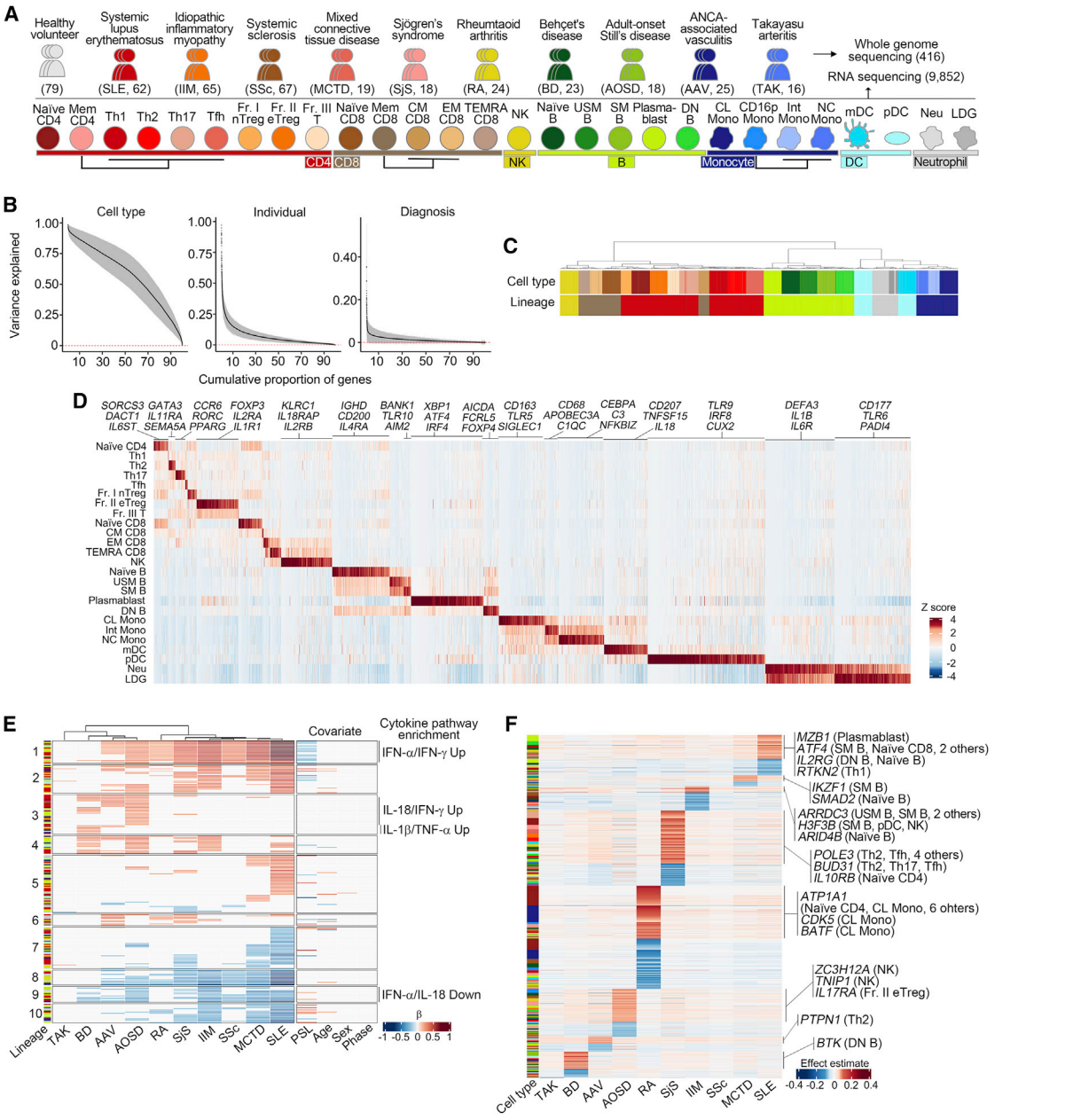


Figure 1. Diversity in gene expression among immune cells and immune-mediated diseases

(A) Summary of the samples collected in this study. Sample numbers after quality control are presented in parentheses. Color coding of each cell type in this article is illustrated, and the rectangles indicate the lineages to which they belong. When both the parent population and its subsets were analyzed, their relationship is indicated beneath with thin lines. A full description of subset names is in Table S8. See also Figure S1 for sorting strategy of immune cell subsets.

(B) Proportion of gene-expression variance explained by cell types, differences among individuals and differences among clinical diagnoses. Gray shading indicates ±1.96 SD.

(C) Hierarchical clustering of RNA-seq samples based on the 5,000 most variable genes. The top labels are colored by each cell type, and the bottom labels are colored by their lineages. The color code is indicated in (A).

(D) Cell-type-specific expressed genes. Column-wise *Z* scores of normalized counts are plotted. Representative genes are annotated on the top. Full lists of cell-type-specific genes are in Table S2.

(E) Comparison of gene-module expression among IMDs. The associations of gene-module eigenvectors and clinical diagnoses were assessed with a linear regression model. β coefficients comparing the patients and healthy volunteers are plotted when the differences are significant (FDR <0.01). Disease names are hierarchically clustered based on the correlation distance of β coefficients. The associations of module eigenvectors and covariates included in the analysis are

(legend continued on next page)

Many eQTL studies in various tissues and cell types have been performed (Chen et al., 2016; GTEx Consortium, 2020; Lappalainen et al., 2013; Schmiedel et al., 2018). The Genotype-Tissue Expression (GTEx) consortium has conducted large-scale tissue eQTL analyses and greatly advanced this field. They have unraveled the tissue or cell-type specificity of eQTLs (Kim-Hellmuth et al., 2020), the impact of sample size on eQTL discovery (GTEx Consortium, 2020), and the importance of analyzing relevant tissue eQTLs for interpretation of GWAS signals (Gamazon et al., 2018). Nonetheless, it is important to note that GTEx has not focused on immune cells, which have been generally included as whole-blood or EBV-transformed lymphocytes. As for immune cell eQTLs, some datasets including BLUEPRINT (Chen et al., 2016) and DICE (Schmiedel et al., 2018) have clarified the genetic functions in immune cells and their cell-type-specific features. However, they were limited in the number of cases (~197 and 91 samples per cell type, respectively) and the comprehensiveness of cell types (3 and 13 cell types, respectively). For example, plasmablasts or dendritic cells, both of which are important for autoimmunity (Ganguly et al., 2013; Martin and Chan, 2004), were not collected in these studies.

An important feature of eQTL analyses in immune cells is that there are dynamic variations in response to external stimuli (Alasoo et al., 2018; Fairfax et al., 2014; Simeonov et al., 2017). For that reason, studies of immune cells from healthy donors might not capture the range of eQTL effects found in disease states. Also, while experimental stimulation of immune cells can mimic disease status to some extent, it does not always reflect the *in vivo* physiologic environment (Tawfik et al., 2020).

We describe here our construction of a database of gene expression as well as genome sequencing data from a large variety of immune cells from IMD patients. Collectively, those data are termed the "Immune Cell Gene Expression Atlas from the University of Tokyo (ImmuNexUT)." We identified immune cell-type-specific eQTL effects, including previously unreported rare cell types, their diversity under physiological inflammatory conditions, and significant association of these eQTLs with IMD genetics. Our non-European, comprehensive, diversified atlas of immune cells will allow researchers to better understand immunogenetic functions *in vivo*.

RESULTS

Diversity in gene expression among immune cells and immune-mediated diseases

After quality control, we examined gene-expression patterns across 9,852 samples from 416 donors. The samples were provided by 79 healthy volunteers and 337 patients who were diagnosed with systemic lupus erythematosus (SLE), idiopathic inflammatory myopathy (IIM), systemic sclerosis (SSc), mixed connective tissue disease (MCTD), Sjögren's syndrome (SjS), rheumatoid arthritis (RA), Behçet's disease (BD), adult-onset Still's disease (AOSD), ANCA-associated vasculitis (AAV), or Takayasu arteritis (TAK) (Figure 1A; Table S1). We purified 28 distinct immune cell types that included almost every type of peripheral immune cells (Figure S1).

With a random-effects model, the largest proportion of gene-expression variance was explained by cell-type differences (Figure 1B). Individual differences also explained non-zero variance of expression for most of the genes, and clinical diagnoses explained variance for some genes to a lesser extent (Figure 1B). Hierarchical clustering demonstrated that gene-expression patterns accurately recapitulated fractionated immune cell subtypes (Figure 1C). Each subset harbored specifically expressed genes, including those for cytokine receptors or pattern recognition receptors, suggesting their distinct responses to environmental signals (Figure 1D; Table S2). Specifically expressed genes included well-known lineage-specific transcription factors (Oestreich and Weinmann, 2012) (e.g., *FOXP3* in regulatory T cells and *RORC* in Th17 cells) as well as previously unregarded ones (e.g., *FOXP4* in double-negative [DN] B cells), indicating that our comprehensive dataset serves as a useful atlas for the study of immune cells.

We constructed data-driven gene modules and compared their expression levels between IMDs and healthy volunteers (STAR Methods). When we compared the patterns of dysregulated gene modules, IMDs were divided into 2 groups, largely corresponding to clinically distinct autoimmune diseases (SLE, MCTD, SSc, SjS, IIM, and RA) and autoinflammatory diseases (BD and AOSD) (Figure 1E). For the annotation of gene modules to biological pathways, we separately constructed gene modules with single-cytokine stimulated synovial fibroblast RNA sequencing (RNA-seq) data (Tsuchiya et al., 2020) and assessed their overlaps (STAR Methods; Figure S2A). Gene modules upregulated in autoimmune diseases (cluster 1 in Figure 1E) showed significant overlap with interferon (IFN)-induced gene sets, whereas gene modules upregulated in autoinflammatory diseases (cluster 3) showed significant overlap with interleukin (IL)-18- or IL-1β-induced gene sets (Figure S2A), consistent with previous disease-by-disease studies (van Kempen et al., 2015). When we compared the expression of these modules between individual donors, autoimmune disease patients, especially those with SSc, IIM, and RA, showed heterogeneous distributions (Figure S2B), consistent with heterogeneity both within and between autoimmune diseases (Cho and Feldman, 2015). When we compared patients with autoantibodies, a significant proportion of IIM patients with antibodies against the melanoma differentiation associated protein 5 (anti-MDA5), known to be associated with severe clinical symptoms (Sontheimer, 2017), belonged to the high interferon signature group (two-sided Fisher's exact test, $p = 0.0048$, Figure S2C). The patient heterogeneity among diseases underpins the usefulness of cross-disease datasets for improved understanding of IMD patients.

plotted on the right. Representative results of cytokine pathway enrichment analysis of gene modules are shown. Full results of enrichment analysis are in Figure S2A.

(F) Genes dysregulated in specific diseases. Effect estimates compared to healthy volunteers are plotted (STAR Methods). Genes are arranged according to the cell types and the effect estimates. Representative upregulated genes are shown on the right. A full list of the genes is in Table S3.

See also Figures S1 and S2.

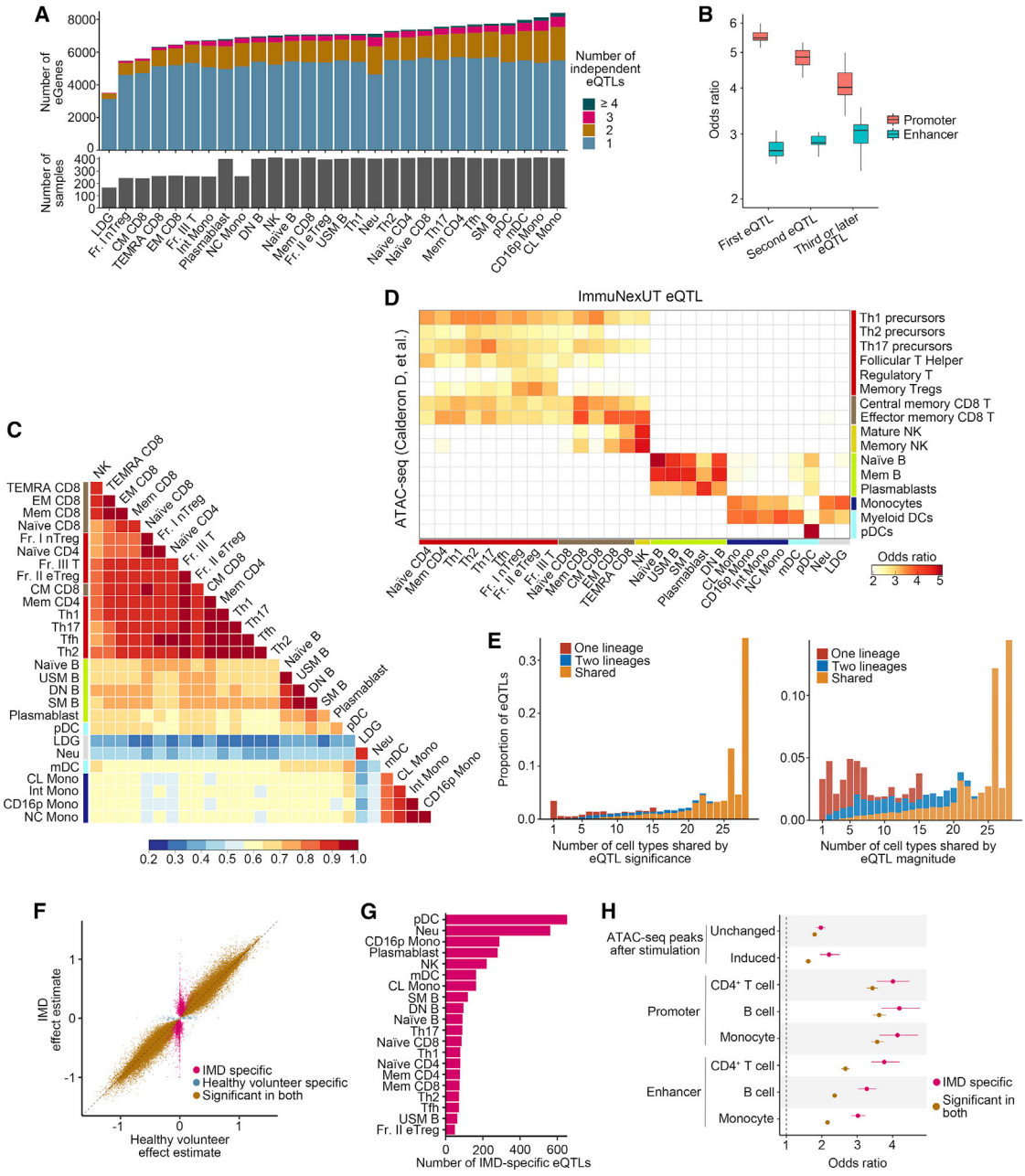


Figure 2. Expression QTLs show immune cell-type specificity and the influence of *in vivo* stimulation

(A) The number of significant eGenes in each cell subset (top) and the number of samples used for eQTL analysis (bottom). Top bars are colored by the number of conditionally independent eQTLs for each gene.

(B) Enrichment of conditionally independent eQTLs in promoters or enhancers of the corresponding cell types from Roadmap data (STAR Methods) stratified by rank.

(C) Pairwise sharing by magnitude of eQTLs among immune cell subsets. For each pair of immune cells, the proportion of eQTLs that have the same sign and the effect size within a factor of 2 of one another is plotted. Cell types are arranged by hierarchical clustering. The colored bars (left) indicate the cell-type lineages.

(D) Overlap of top eVariants (x axis) with differentially accessible regions from independent ATAC-seq study (y axis). Subset names of ATAC-seq data are from the original article (Calderon et al., 2019). The color bars indicate the cell-type lineages.

(E) Distribution of the number of cell types in which eQTLs were shared by significance (left) or magnitude (right). QTL magnitude was defined to be shared when the effect estimate outputted by mash was within a factor of 2. QTL was defined to be significant if it had a mash local false sign rate (LFSR) of <5%. Here, each cell type is classified into 1 of 3 lineages: T/NK cells, B cells, and myeloid cells. The bar chart is colored according to their sharing by the 3 lineages.

(legend continued on next page)

Please cite this article in press as: Ota et al., Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases, Cell (2021), https://doi.org/10.1016/j.cell.2021.03.056

Along with the pathways commonly dysregulated in IMDs, some genes were especially dysregulated in specific diseases (Figures 1F and S2D; Table S3). These included previously indicated genes (e.g., *MZB1* and *ATF4,* possibly associated with excessive endoplasmic reticulum stress, upregulated in SLE B cells [Miyagawa-Hayashino et al., 2018; Zhang et al., 2015]). However, the majority of the genes were previously unreported (e.g., *ZC3H12A* and *TNIP1*, associated with the modulation of the nuclear factor κB (NF-κB) pathway [G'Sell et al., 2015; Matsushita et al., 2009], upregulated in AOSD natural killer [NK] cells). Thus, our resource is consistent with well-characterized features of IMDs and might be useful in investigating previously unrecognized pathogenesis of IMDs in the future.

Expression QTLs show immune cell-type specificity

To characterize the regulation of gene expression in immune cells, we performed eQTL analysis with deep whole-genome sequence (WGS) data with a mean coverage of 41×. Our genome data showed good concordance of allele frequency with East Asian samples from the 1000 Genomes Project (Auton et al., 2015) and the Japanese WGS reference panel (Tadaka et al., 2018) (Figures S3A–S3C). In each cell type, we performed a forward-backward stepwise linear regression (Battle et al., 2017) between autosomal expressed genes (eGenes) and genetic variants (eVariants) located within 1 Mb of the transcription start site. Covariates included clinical diagnoses and latent factors. We empirically adjusted for the number of latent factors according to sample sizes following previous reports (GTEx Consortium, 2020) (Figures S3D and S3E; STAR Methods). Hereafter, we denote independent eVariants with the highest association with eGene in each regression trial as the "top eVariants."

Our analysis identified eQTLs for 13,395 protein coding genes and 3,839 long non-coding RNAs at a 5% false discovery rate (FDR). A median of 7,092 genes were identified as eGenes in each cell type (Figure 2A), which is 2.2-fold more than that identified in the DICE database (Schmiedel et al., 2018). Consequently, the majority of eGenes was shared among cell types (Figure S3F), which is not concordant with the information that DICE reported. This discrepancy was not solely due to small subdivision of immune cells in our dataset because eGenes were shared between lineages even when we categorized immune cells into 3 large lineage groups (T/NK cells, B cells, and myeloid cells; Figure S3F). Expression QTL effect sizes were significantly higher in eGenes that were detected in both of the datasets than eGenes that were detected only in ImmuNexUT ($p < 2.2 \times 10^{-16}$, Mann-Whitney U test). Taken together, we surmise that, while DICE identified cell types in which eQTLs had relatively large effect sizes, the other cell types also harbored eQTLs for the same genes but with smaller effect sizes, effects that were detected in ImmuNexUT.

We identified 2 or more conditionally independent eVariants (i.e., variants independently associated with gene expression in conditional analysis) in 25% (median) of eGenes (Figure 2A). Compared to the primary eVariants, these independent second or later eVariants showed greater enrichment in enhancers of matching Roadmap cell types as previously reported (Kundaje et al., 2015; Battle et al., 2017; Ishigaki et al., 2017) (Figure 2B). When eQTLs were stratified by effect sizes, low effect size eQTLs overlapped with enhancers as well as high effect size eQTLs, on the contrary to the predominant overlap of promoters with high effect size eQTLs (Figure S3G). As cell-type-specific gene expression is largely regulated by enhancers (Heinz et al., 2015), those results indicate the importance of large sample size in identifying cell-type-specific eQTLs, which frequently have low effect sizes. The following observations from our dataset confirmed the cell-type specificity of eQTL effects. First, by comparing the sharing of eQTL effect sizes among cell types with the multivariate adaptive shrinkage (mash) method (Urbut et al., 2019), we observed an obvious pattern of similarity among biologically related cell subsets (Figure 2C). Second, when we compared our top eVariants with open chromatin regions using data from assay for transposase-accessible chromatin sequencing (ATAC-seq) (Calderon et al., 2019), eVariants showed high overlap with differentially accessible regions of the respective cell type (Figure 2D). Also, our top eVariants were enriched in enhancer and promoter elements of Roadmap immune cells (Kundaje et al., 2015), with higher cell-type specificity for enhancers of relevant cells (Figures S3H–S3J). Those data suggested that immune cell eQTLs are in part driven by cell-type-specific elements. When we compared local false discovery rate (LFSR) outputted by mash (Urbut et al., 2019; STAR Methods), which is a metrics for eQTL significance, more than half of eQTLs had detectable effects in majority of immune cell types (>25 out of 28, Figure 2E, left panel). On the contrary, when focusing on the effect estimates outputted by mash, many eQTLs had variable effect sizes across cell types, as well as lineages (Figure 2E, right panel). These observations are in accordance with a previous report from a tissue eQTL study (Urbut et al., 2019).

Immune-mediated disease-specific eQTLs demonstrate the influence of *in vivo* stimulation

To assess the impact of including patient-derived samples for eQTL identification, we next separately performed eQTL analysis in each cell subset only with healthy volunteer samples or only with IMD patient samples. For a fair comparison, here we limited the analyses to variants with minor allele frequency >0.1 in both groups and genes expressed in both. We combined the eQTL effects of top eVariants with mash. As expected, majority of eQTLs were shared (94% of eQTLs attained LFSR <0.05 in both) although the absolute values of the effect estimates were significantly higher in IMD patients on average ($p < 2.2 \times 10^{-16}$, Mann-Whitney U test). Some eQTLs were significant only in IMD patients (Figure 2F). The number of IMD-specific eQTLs varied

(F) Comparison of effect estimates of top eVariants between healthy volunteers and IMD patients. Results of 20 cell types are plotted together (STAR Methods).
(G) The number of IMD-specific eQTLs in each cell subset.
(H) Enrichment of IMD-specific eQTLs and eQTLs significant in both healthy volunteers and IMD patients to functional annotations (STAR Methods). Bars indicate 95% CI.
See also Figure S3.

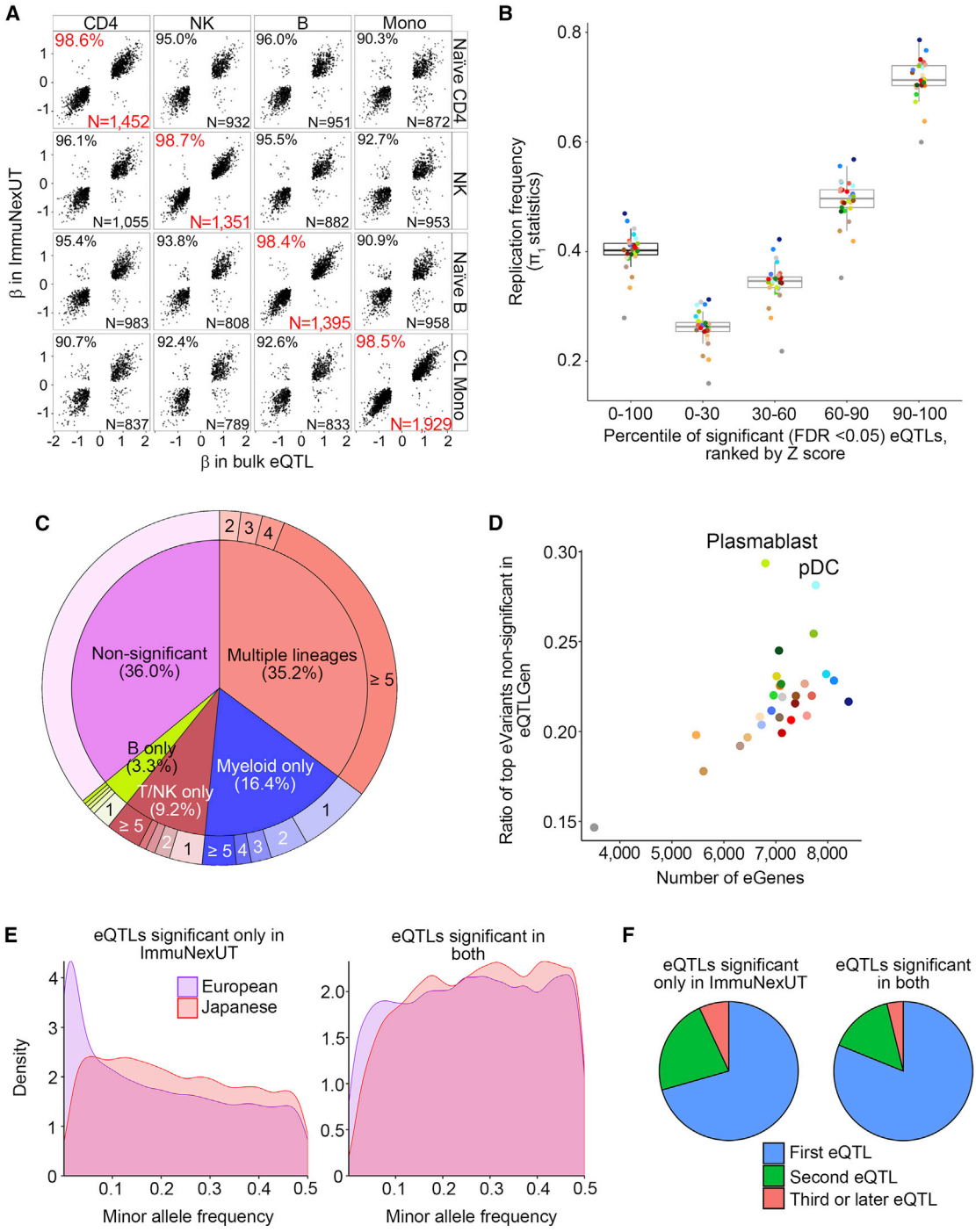


Figure 3. ImmuNexUT exhibits good concordance with previous eQTL studies and provides complementary information

(A) Comparison of β coefficients between ImmuNexUT eQTLs and previously reported bulk immune cell eQTLs (Ishigaki et al., 2017). Ratio of eQTLs in a consistent direction and the number of eQTLs significant in both studies are annotated at top left and right bottom of each plot, while highlighted in red for pairs of corresponding cell types.

(B–F) Comparisons with large-scaled whole-blood eQTL study (Võsa et al., 2018) (eQTLGen, below).

(B) Replication frequency (π_1 statistics) of significant cis-eQTLs discovered in eQTLGen. Significant eQTLs (FDR <0.05) in eQTLGen are further divided into each bin according to percentiles of absolute values of *Z* scores, and their replication frequency in ImmuNexUT are evaluated in each cell type. The color of each dot indicates cell type as illustrated in Figure 1A.

(C) Proportions of top eQTLs from eQTLGen that were significant (FDR <0.05) in each cell-type analysis of ImmuNexUT. Cells were divided into 3 lineages: T/NK cells, B cells, and myeloid cells. We compared the lineage sharing (pie chart) and the number of significant cell types in each category (outer ring).

(legend continued on next page)

Please cite this article in press as: Ota et al., Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases, Cell (2021), https://doi.org/10.1016/j.cell.2021.03.056

among cell types, with a tendency for larger number in myeloid cells (Figure 2G). Compared to eQTLs significant in both healthy volunteers and IMD patients, IMD-specific eQTLs were significantly more enriched in enhancers and in immune cell ATAC-seq peaks that were induced after stimulation (Calderon et al., 2019) (Figure 2H). When assessing the overlap with disease-associated variants, IMD-specific eQTLs, as well as eQTLs shared with healthy volunteers, showed significant enrichment in IMD-associated GWAS variants (Figure S3K). These observations suggested that using patient-derived samples for eQTL identification helps to identify stimulation-dependent eQTLs under physiologic conditions, which are relevant to disease biology and difficult to detect with only healthy volunteer samples.

ImmuNexUT exhibits good concordance with previous eQTL studies and provides complementary information

To evaluate the validity of our eQTLs, we compared our results with 2 external datasets. First, we compared our data with our previously reported bulk immune cell eQTL study of 105 healthy Japanese volunteers with no sample overlap (Ishigaki et al., 2017). There was good concordance in eQTL effects (Figure 3A, binomial $p < 2.2 \times 10^{-16}$ for each pair of corresponding cell types), supporting the reproducibility of our data.

Next, we compared our data with the largest current whole-blood cis-eQTL dataset based on the analysis of >30,000 Europeans (Võsa et al., 2018) (eQTLGen). For the significant eQTLs from eQTLGen (FDR <0.05), a median of 40% was replicated in each immune cell, and this proportion became much higher with the more stringent eQTL significance thresholds in eQTLGen (Figure 3B). Of the best eVariants per eGene in eQTLGen that were polymorphic in Japanese donors, 64% were also significant in at least 1 immune cell type (Figure 3C). By comparison, 20%–25% of top eVariants in our dataset that were polymorphic in Europeans were not significant in eQTLGen (Figure 3D). Although this ratio was largely proportional to the number of eGenes, it was disproportionally high in plasmablasts and plasmacytoid dendritic cells (pDCs) (Figure 3D). This outcome might be due to the scarcity and relatively distinct eQTL features of these cell types (Figures 2C and S1). Also, ImmuNexUT top eVariants that were not significant in eQTLGen were biased to lower allele frequencies in Europeans and second or later conditionally independent eQTLs (Figures 3E and 3F), which might be difficult to detect in even large-scale whole-blood data.

Combining eQTL datasets from different populations improves accuracy for fine mapping

In order to evaluate the utility of our Japanese eQTL data for fine mapping of causal eQTL variants, we jointly fine-mapped our data with European whole-blood data from GTEx v8 (GTEx Consortium, 2020), in which WGS analysis was also performed. We conducted this analysis with classic monocyte and neutrophil data that have relatively large overlaps of eQTLs with whole blood. Compared to the independent analysis, the joint analyses significantly reduced the number of candidate causal variants (95% credible set: the variant set that was assumed to include all the causal variants with 95% confidence) while maintaining a strong enrichment in experimentally validated functional variants by reporter assay (van Arensbergen et al., 2019) (Figure 4A). As expected, differences in linkage disequilibrium (LD) likely contributed to narrowing down causal variants in some loci. For example, in the *MFN2* locus (Figure 4B), the eQTL variant was in high LD with many proxy variants in Europeans and it was difficult to fine-map causal eQTL variant with European data. The different LD structure in Japanese subjects aided fine mapping and narrowed down to 5 variants in the 95% credible set with joint analysis, and the top prioritized variant, rs873458, had functional activity in a previously published reporter assay (van Arensbergen et al., 2019). In the *CEP19* locus (Figure 4C), we could narrow down to 3 variants in the 95% credible set with joint analysis, and top prioritized variant, rs71323742, had functional activity in a reporter assay (van Arensbergen et al., 2019). These results suggested the utility of combining datasets from different populations for fine mapping.

Identification of biological pathways that diversify eQTL effects *in vivo*

Based on previous reports regarding the heterogeneous effects of eQTLs either *in vitro* (Fairfax et al., 2014) or *in vivo* (Davenport et al., 2018; van der Wijst et al., 2018; Zhernakova et al., 2017), we analyzed the interaction of eQTL effect size and genome-wide gene expression that are dependent on intrinsic or extrinsic factors (context-dependent eQTL) (Figure S4A) (Zhernakova et al., 2017). Here, genome-wide gene expression was treated as the proxy of various environmental contexts, and we call them "proxy genes" (pGenes) as previously proposed (Zhernakova et al., 2017).

We analyzed the top eVariants in each cell type, adjusted p values by permutations of pGene labels, and identified 37,875 significant pGene-eQTL interactions for 5.6% of eGenes (FDR <0.05, Table S4). The pGenes with significant positive interactions with eQTLs formed distinct clusters based on their interactions (Figure 5A), indicating functionally relevant gene sets associated with similar gene regulatory machinery. As expected, genes that significantly overlapped with curated IFN signature gene set (one-sided Fisher's exact test, $p = 9.2 \times 10^{-111}$) (Liberzon et al., 2015), which can be considered as proxies of IFN activity, made a cluster (P1 in Figure 5A). Gene transcription downstream of IFN is known to be mediated by complexes that consist of transcription factors STAT1, STAT2, and IRF9 (Michalska et al., 2018). Consistently, *STAT1*, *STAT2*, and *IRF9* belonged to cluster P1, and eQTL variants interacting with P1 genes (cluster E1) were significantly enriched in STAT1- or STAT2-binding motif breaking sites

(D) The ratio of ImmuNexUT significant eQTLs that were also polymorphic in Europeans (minor allele frequency >0.1) but not significant in eQTLGen. Data are plotted against the number of eGenes in ImmuNexUT. The color of each dot indicates cell type as illustrated in Figure 1A.
(E) Distribution of minor allele frequencies of eVariants in Europeans and Japanese. Left panel shows for eQTLs significant only in ImmuNexUT. Right panel shows for eQTLs significant in both datasets.
(F) Distribution of ranks of conditionally independent eQTLs significant only in ImmuNexUT (left) and significant in both datasets (right). Variants with minor allele frequency >0.1 in Europeans were utilized for comparison.

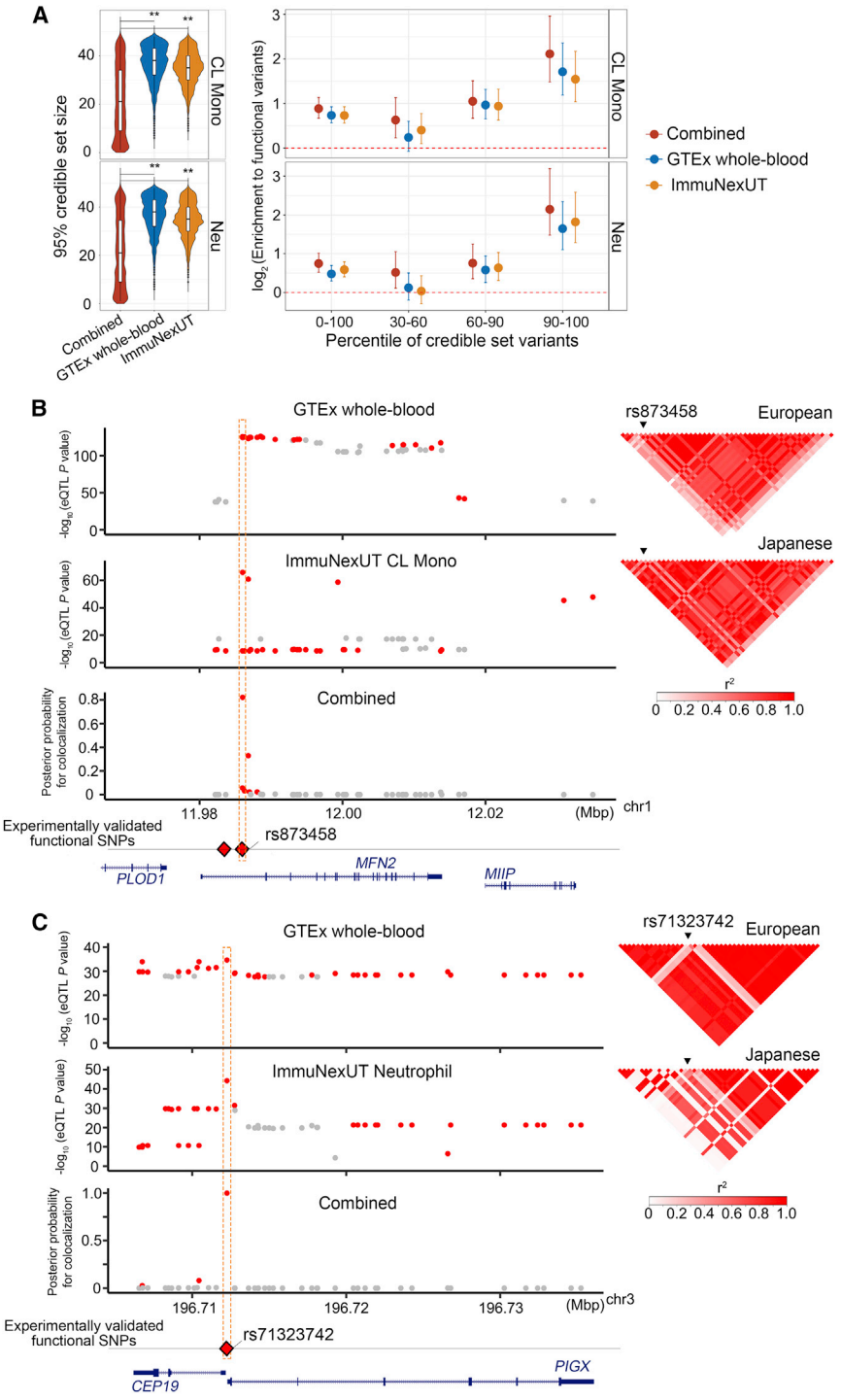


Figure 4. Combining eQTL datasets from different populations improves accuracy for fine mapping

(A) The number of variants in fine-mapped 95% credible sets (left) and the enrichment of fine-mapped variants in experimentally validated functional variants (right). GTEx whole-blood and ImmuNexUT classic monocytes or neutrophils are analyzed independently and jointly. **Two-sided Wilcoxon signed-rank tests, $p < 2.2 \times 10^{-16}$. In the right panel, variants in 95% credible sets were further divided into each bin according to percentiles of posterior probabilities of being causal, and their enrichment to experimentally validated functional variants (van Arensbergen et al., 2019) are plotted. Bars indicate 95% CI.

(B and C) Fine mapping of *MFN2* eQTL variants (B) and *CEP19* eQTL variants (C). In the left panel, variants in 95% credible sets in independent and joint analysis are highlighted with red. Variants with functional activity in a reporter assay (van Arensbergen et al., 2019) are highlighted at the bottom. In the right panel, r^2 measures of linkage disequilibrium between analyzed variants are plotted in European and Japanese population.

($p = 5.9 \times 10^{-9}$), supporting the validity of these interactions. Cluster P3 showed significant overlap with cell-cycle-associated genes ($p = 9.6 \times 10^{-37}$), which might be proxies of cell proliferation (Whitfield et al., 2006). Cluster P8 showed significant association with eQTLs in DN B cells (cluster E7) and showed significant overlap with upregulated genes in DN2 B cells (Jenks et al., 2018) ($p = 1.6 \times 10^{-19}$). Thus, this cluster might be a proxy of the DN2 B cell population. High IFN activity in autoimmune patients (represented by *STAT1* [Figure 5B]), expansion of DN2 B cells in SLE patients (represented by *FCRL3* [Figure 5C]), which is consistent with the previous report (Jenks et al., 2018), and various cell-proliferation patterns among IMD patients (represented by *CDCA7* [Figure 5D]) were associated with variations of these events. Additionally, cluster P4 contained genes that were significantly downregulated with aging ($p = 5.3 \times 10^{-14}$; STAR Methods). As P4 contained *SATB1* (Figure 5E), whose downregulation is associated with cellular senescence (Satoh et al., 2013), this cluster might be a proxy of cellular aging and/or senescence.

Overall, the pGenes, especially those with many interactions with eQTL effects, showed larger variation among IMDs compared with the rest of the genes (Figure 5F). The pGenes were enriched in differentially expressed genes (DEGs) of IMDs (Figure S4B). Also, the number of context-dependent eQTLs significantly decreased when disease variance was regressed from genome-wide gene-expression data (Figure S4C). These results indicated that including patient-derived samples improved power for the detection of context-dependent eQTLs by increasing variance in expression of the pGenes. Context-dependent eGenes showed significant overlap with genes

Please cite this article in press as: Ota et al., Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases, Cell (2021), https://doi.org/10.1016/j.cell.2021.03.056

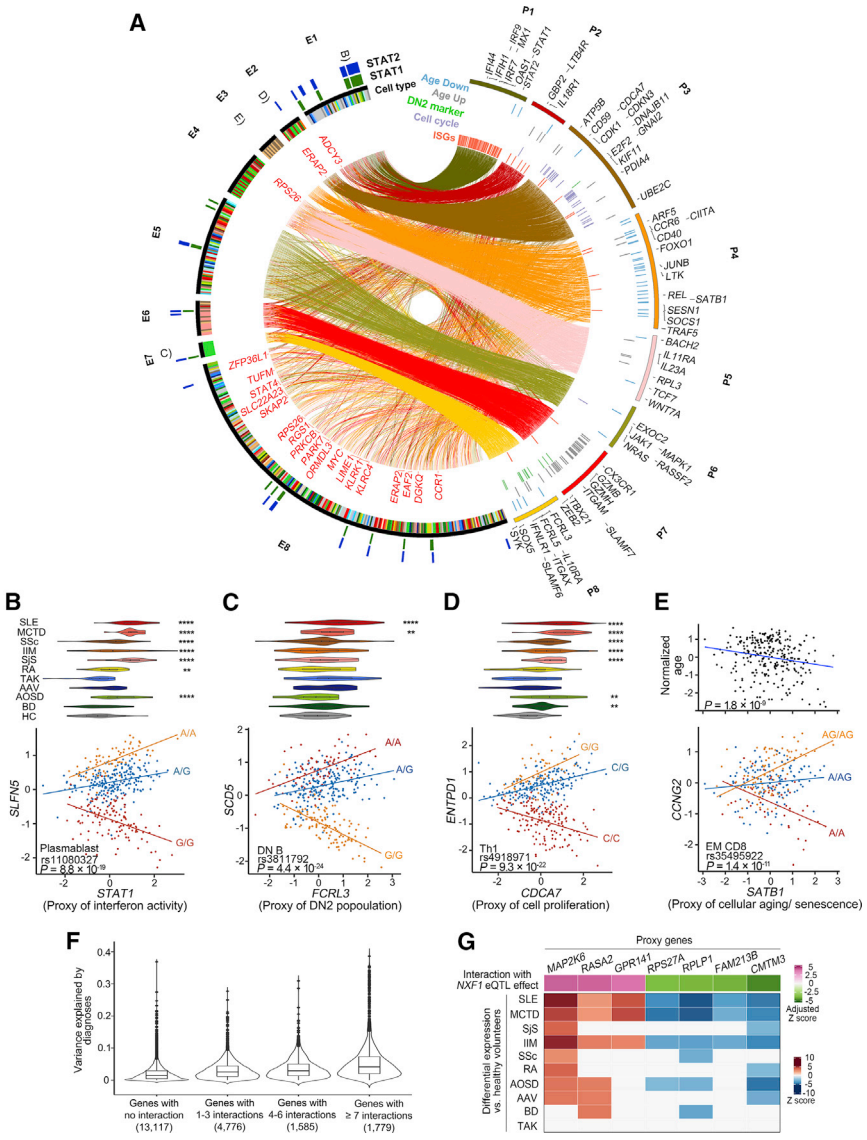


Figure 5. Identification of biological pathways that diversify eQTL effects *in vivo*

(A) Context-dependent eQTLs and their associated events. For proxy genes (pGenes; right part), representative gene names are depicted with black characters, and genes corresponding to described gene sets are labeled with color in inner rings. For eQTLs (left part), variants with STAT1 or STAT2 binding motif breaking effects and those illustrated in (B)–(E) are marked. Red characters indicate eGenes whose eVariants are in strong LD with IMD GWAS top variants. To facilitate interpretation, eQTLs and pGenes were filtered according to the strength and the number of interactions, and positive and strong pGene-eQTL interactions are linked (STAR Methods). ISG, interferon signature gene. DN2, double-negative 2 B cell.

(B–D) Representative interaction plots of pGene (x axis) and eGene (y axis) expression. P values were calculated by the interaction analyses of pGene expression and eQTL effect sizes (STAR Methods). Top violin plots compare the expression of pGenes among clinical diagnoses. Two-sided Welch's t test. **p < 0.01; ***p < 0.001; ****p < 0.0001 compared to healthy controls (HCs).

(E) Interaction plot of *SATB1* expression and *CCNG2* eQTL effects. In the top plot, biological age was log-normalized and compared with *SATB1* expression.

(F) Proportion of gene-expression variance explained by diagnoses (10 IMDs or healthy state), while genes are stratified by the number of genome-wide significant interactions with eQTL effects. The numbers in parentheses indicate the number of genes in each group.

(G) Interaction of *NXF1* (rs11304762) eQTL effect and pGenes in classic monocytes and differential expression of pGenes in classic monocytes. All pGenes with FDR <0.05 are shown, and differential expression *Z* scores are plotted when the differences are significant (FDR < 0.05).

See also Figure S4.

induced after vaccination ($p = 2.4 \times 10^{-17}$), viral infection ($p = 9.8 \times 10^{-8}$), or LPS stimulation ($p = 3.6 \times 10^{-7}$). When we stratified context-dependent eGenes by their association with disease variance, these enrichments were higher in disease variance-associated eGenes than others (Figure S4D; STAR Methods). Thus, context-dependent eQTLs of immune cells, especially those associated with disease conditions, may play a role in individual differences in inflammatory response or immune response. For instance, *NXF1*, a regulator of TLR7-driven IRF5 transcriptional activity (Fu et al., 2017), receives context-dependent eQTL effect in classic monocytes (Figure 5G). The pGenes included *MAP2K6* and *RASA2*, members of mitogen-activated protein kinase (MAPK) and RAS activator, that were differentially expressed in IMDs in direction concordant with the eQTL interaction. Thus, this eVariant may diversify individuals' response under MAPK signaling, which is activated in IMD patients, via modifying the expression of *NXF1*.

ImmuNexUT links IMD genetics to cell types, genes, and environment

We then utilized our eQTL dataset for interpretation of IMD-associated GWAS signals. We assessed the association of our eQTLs with GWAS results using stratified LD score regression (Finucane et al., 2015) (STAR Methods). As expected, most of the eQTL annotations were significantly associated with the heritability of a variety of complex traits and diseases, possibly reflecting eQTL elements shared by broad range of cell types and tissues (Hormozdiari et al., 2018) (Figure S4E). When we conditioned the shared elements by jointly regressing the meta-analyzed eQTL annotation (STAR Methods), most of the associations of non-immune traits diminished, although specific associations of immune diseases and immune cell eQTLs remained (Figure 6A). These results support the cell-type specificity of heritability enrichment. Our eQTL annotations have variant-level cell-type-specific functional information that is

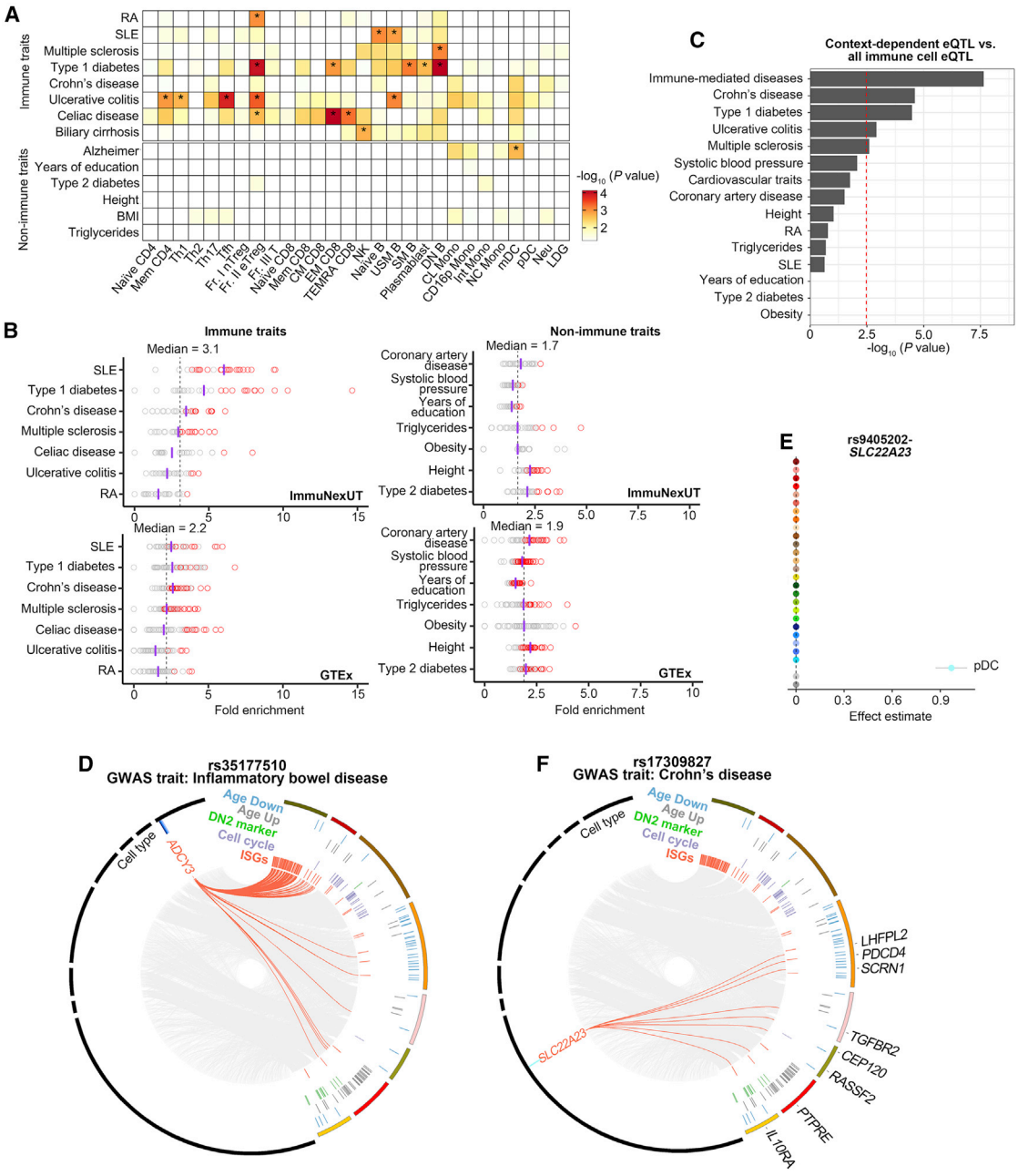


Figure 6. ImmuNexUT links IMD genetics to cell types, genes, and environment

(A) Association significance of each cell-type-specific eQTL annotation and polygenic GWAS signals in stratified LD score regression analysis, after multiple regression with meta-analysis annotation (STAR Methods). *FDR <0.05.

(B) Fold enrichment of ImmuNexUT or GTEx eQTL variants in GWAS top variants. Traits with at least 30 genome-wide significant loci were analyzed. Red circles indicate tissue-trait pairs that passed Bonferroni significance threshold. Purple bars and dashed lines indicate the median values of the fold enrichment in each trait and all traits in the category.

(C) Enrichment of context-dependent eQTLs in GWAS top signals compared with all immune cell eQTLs. Red dotted line is the cutoff for Bonferroni significance.

(D) Context-dependent eQTLs that are in tight LD with inflammatory bowel disease GWAS top variant rs35177510 ($r^2 \geq 0.8$ in EUR and EAS) and their association with pGenes. The pGenes and eQTLs are filtered and arranged in the same order as Figure 5A. Interactions with adjusted *Z* score >2.5 are linked with red lines.

(E) Comparison of eQTL effect estimates of rs9405202, which is in perfect LD with Crohn disease GWAS top variant rs17309827 ($r^2 = 1$ in EUR and EAS), to *SLC22A23* outputted by mash. The gray bar indicates ±1.96 SD.

(F) Context-dependent eQTL which is in perfect LD with Crohn disease GWAS top variant rs17309827 and its association with pGenes. The pGenes and eQTLs are filtered and arranged in the same order as Figure 5A. Interactions with adjusted *Z* score >2.5 are linked with red lines. The interaction with *TGFBR2* attained genome-wide significance ($p = 3.3 \times 10^{-10}$).

See also Figure S4.

Please cite this article in press as: Ota et al., Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases, Cell (2021), https://doi.org/10.1016/j.cell.2021.03.056

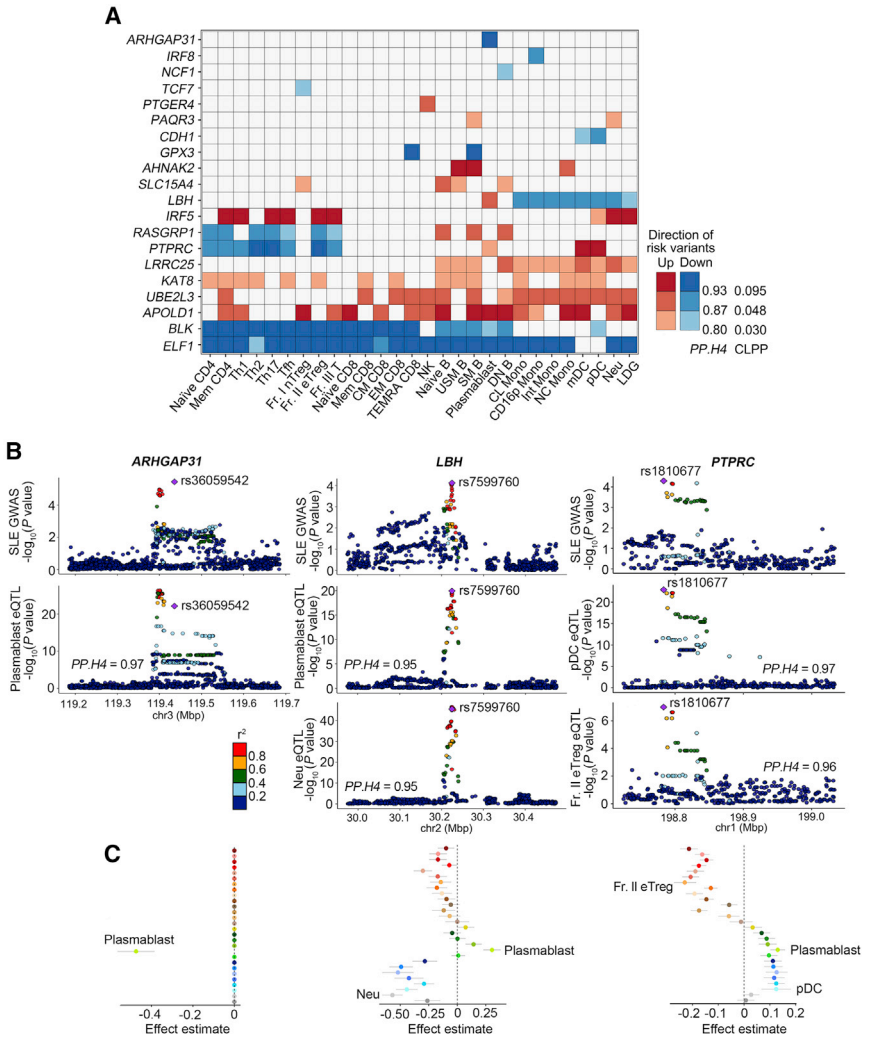


Figure 7. SLE risk genes prioritized by co-localization analysis

(A) Colocalization of Japanese SLE GWAS variants and eQTLs. Posterior probabilities of colocalization are signed with the directional effect of risk variants. As metrics for colocalization probability, CLPP from eCAVIAR is plotted for 2 genes (*CDH1* and *GPX3*) while *PP.H4* from coloc is plotted for the other genes (STAR Methods).

(B) Regional association plots of SLE GWAS and immune cell eQTLs for *ARHGAP31* (left), *LBH* (middle), and *PTPRC* (right) loci.

(C) Comparison of eQTL effect estimates of the top eQTL variants which are in LD with GWAS top variants in Japanese in *ARHGAP31* (left, rs62266700), *LBH* (middle, rs12714303), and *PTPRC* (right, rs10919581) loci outputted by mash. Gray bars indicate ±1.96 SD.

See also Figure S4.

qualitatively different from region-level epigenome annotation. We reconfirmed previous findings with epigenome annotation including the association of regulatory T cells (Fr. II eTreg) with RA (Kundaje et al., 2015; Trynka et al., 2013) and type 1 diabetes (Kundaje et al., 2015) and the association of B cells with SLE (Kundaje et al., 2015). In addition, we identified the associations of DN B cells with multiple sclerosis and type 1 diabetes, $CD8^+$ T cells with celiac disease, and NK cells with biliary cirrhosis (Figure 6A).

In some cases, the eQTL effects of GWAS top signals point to disease susceptibility genes (Gamazon et al., 2018). Therefore, we assessed the enrichment of top variants from the NHGRI-EBI GWAS catalog (Buniello et al., 2019) in our top eVariants by proximity (Figure 6B). All of the eVariants in LD with IMD GWAS top variants, corresponding eGenes and cell types, are listed in Table S5. Compared with GTEx v8 data (GTEx Consortium, 2020), there was increased enrichment of immune trait GWAS signals in our immune cell eQTLs (e.g., SLE GWAS top variant enrichment to B cells eQTL; Figure 6B; Table S6). Those results contrasted with the non-immune trait GWAS enrichment in GTEx eQTLs (e.g., coronary artery disease GWAS enrichment to coronary artery eQTL; Figure 6B; Table S6).

It was notable that IMD GWAS signals showed further enrichment in context-dependent eQTLs compared to all immune cell eQTLs ($p = 2.3 \times 10^{-8}$; odds ratio = 10.8, Figure 6C). For example, an IBD-associated variant showed an IFN-dependent eQTL effect for *ADCY3* in monocytes (Figure 6D). Moreover, a Crohn-disease-associated variant had an eQTL effect for *SLC22A23* that was specific in pDCs (Figure 6E) and this eQTL effect interacted with some genes including *TGFBR2* (Figure 6F) that codes for a component of the transforming growth factor (TGF)-β receptor. Dendritic cell-specific knockout of *TGFBR2* in mice results in autoimmune phenotypes including colitis (Ramalingam et al., 2012), supporting the cell-type-specific disease relevance of this pathway. Considering IFNs and TGF-β play key roles in IBD pathogenesis (Friedrich et al., 2019), these interactions illustrate modulation of GWAS variant functions by disease-associated environmental factors.

Next, based on the significant overlap of immune cell eQTLs with SLE GWAS top signals (Figure 6B), we prioritized SLE risk genes by assessing colocalization of GWAS signals and eQTL signals. Of the 32 suggested associated loci outside the HLA region in recently reported Japanese SLE GWAS (Akizuki et al., 2019), 3 were in strong LD with missense variants (STAR Methods; Table S7). Among the remaining 29 loci, 20 showed colocalization with at least one immune cell eQTL with stringent criteria (STAR Methods; Figure 7A; Table S7). Of those, some showed subset-specific or directionally opposite eQTL effects among immune cells. The *ARHGAP31* eQTL effect was observed only in plasmablasts and showed strong colocalization with a GWAS signal (Figures 7B and 7C). Subset specificity was further supported by data that showed that the open chromatin

Please cite this article in press as: Ota et al., Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases, Cell (2021), https://doi.org/10.1016/j.cell.2021.03.056

status of this locus occurred only in plasmablasts among various immune cell types (Figure S4F). Intersecting variants that were in LD with the top eVariant and the GWAS top variant, and that overlapped with open chromatin regions enabled us to narrow down putative causal variants, reinforcing the utility of combining multi-omics datasets (Figure S4F). *LBH* showed not only downregulation in myeloid cells but also upregulation in plasmablasts by the risk allele, and both effects showed nearly perfect colocalization with the GWAS signal (Figures 7B and 7C). As *LBH* is associated with cell expansion and maintenance (Lindley et al., 2015), its high expression in plasmablasts might contribute to its expansion, whereas its downregulation could lead to DNA damage in other cells (Matsuda et al., 2017). *PTPRC*, which codes for the well-known signaling molecule CD45, also exhibited opposing regulatory functions among immune cells by a SLE risk allele (Figures 7B and 7C). These subset-specific and opposing eQTLs likely have a great impact on immune cell orchestration and can be relevant to complex disease pathogenesis.

DISCUSSION

This report constitutes an initial description of our atlas of immune cell gene regulation. Our dataset has distinct and advantageous characteristics compared to previously reported resources, including comprehensive immune cell subdivision, variation of immunological conditions among donors and homogeneous non-European donors.

In gene-expression analysis, our comparison of multiple IMDs in a single platform permitted us to characterize each disease (patient) and identify genes that were associated specifically with a single disease, as well as pathways that were dysregulated in a group of IMDs. For example, AAV, whose etiology is largely unknown but involvement of both of the innate and adaptive immunity has been inferred (Kitching et al., 2020), showed upregulation of IL-18-inducible pathways, and the gene-expression pattern was similar with that of autoinflammatory diseases. This result suggests the potential of repositioning drugs for autoinflammatory diseases as a treatment of AAV and vice versa, although patient heterogeneity should be considered.

Our dataset enabled us to detect physiological (disease) context-driven perturbations of eQTL effects in each specific cell type. Some of the context-dependent eQTLs seemed to track with disease severity (e.g., IFN signal strength and MAPK signal strength). The eVariants interacting with these disease-associated pathways would contribute not only to disease susceptibility but also to the heterogeneity in inflammatory response among patients and should be candidates for future studies about the phenotypic variance within and between IMDs.

We observed an association of IMD-associated genetic variants and immune cell-type-specific eQTLs. Colocalization analysis identified that 63% (20/32) of SLE GWAS top variants from Japanese population colocalized with immune cell eQTLs. This proportion is higher than that of a previous report in which colocalization of autoimmune disease GWAS and eQTL was estimated to be ~25% using three types of immune cells (Chun et al., 2017). We surmise that our cell-type-specific eQTLs with comprehensive immune cell dataset (28 cell types) contributed to better mapping of GWAS signals. Also, we observed disease-relevant eQTLs with cell-type-specific effects, directionally opposite effects among cell types, and context-dependent effects in specific cell types. These eQTLs were unable to be detected without purifying immune cells, including relatively rare cell types (e.g., plasmablasts or pDCs).

Finally, most of the large-scale eQTL datasets have been generated from European donors (GTEx Consortium, 2020; Chen et al., 2016; Fairfax et al., 2014; Lappalainen et al., 2013; Nédélec et al., 2016; Quach et al., 2016; Schmiedel et al., 2018). Our eQTL dataset of Japanese population helped us to identify eQTL variants in East Asian populations that had not been detected in European populations due to low allele frequency, as well as fine-mapping causal eQTL variants by taking advantage of difference in haplotype architecture. Thus, our dataset serves as a good resource in improving the functional annotation of genomic variations.

Limitations of the study

Although our study provides a comprehensive atlas of gene regulation in immune cells *in vivo*, there are some limitations. First, the patient cohort size is different among different diseases; thus, gene-expression or regulation changes in diseases with small sample sizes might be relatively underestimated compared to diseases with larger sample sizes. Second, because this is an observational study, clinical heterogeneity, especially among medications including glucocorticoids and other immunosuppressants, exists and may be affecting gene expression or regulation in part. Third, due to female-biased susceptibility of IMDs, a large proportion of donors in our dataset consist of female subjects, which is unlike other large-scale eQTL datasets and should be interpreted with caution in a comparative study.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Inclusion and exclusion criteria
  - Sample collection and processing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - RNA sequencing
  - QC of gene expression data
  - Whole genome sequencing and variant calling
  - Validation of WGS calls
  - Analyzed samples
  - Variance decomposition
  - Hierarchical clustering
  - Cell type-specific genes
  - Transcriptome signature construction

- Cytokine pathway enrichment analysis
- Genes especially dysregulated in specific diseases
- Data normalization and eQTL analysis
- Epigenome mark enrichment analysis
- Comparison of eQTL effect sizes
- Comparison of eQTL effects between IMD patients and healthy volunteers
- Comparison with other data
- Fine-mapping of eQTLs
- Context-dependent eQTL analysis
- LD score regression analysis
- Comparison with GWAS catalog variants
- Colocalization analysis of eQTL and SLE GWAS
- Statistical test

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2021.03.056.

ACKNOWLEDGMENTS

The super-computing resource was provided by Human Genome Center, Institute of Medical Sciences, The University of Tokyo (http://sc.hgc.jp/shirokane.html). This study was supported by Chugai Pharmaceutical Co., Ltd., Tokyo, Japan; the Center of Innovation Program from Japan Science and Technology Agency (JST) (JPMJCE1304); and the Japan Agency for Medical Research and Development (AMED) (JP17ek0109103, 19ek0410047, and JP20ek0410074).

AUTHOR CONTRIBUTIONS

M. Ota conducted bioinformatics analysis with the help of Y.N., H.H., K.I., C.T., Y.K., and Y.O. H.H., Y.N., Y. Takeshima, H. Yanaoka, M. Okubo, S.K., Y.S., H. Shirai, J.M., R.Y., S.Y., M.N., Y.I., S.S., H. Shoda, N.B., N.Y., K.M., and H.M. managed and contributed to sample collection, cell sorting, RNA sequencing, and whole-genome sequencing. H. Tsuchiya contributed to FLS data management. H.H., N.B., N.Y., K.M., and H.M. contributed to QC of the RNA sequencing data. S.O., M. Okazaki, H. Tsunoda, T.O., K. Yamamoto, and K.F. designed and managed the project. C.T., S.A., K.O., H. Yoshifuji, and T.M. managed SLE GWAS data. K. Yoshida, D.K., M. Okada, H.K., and K.S. managed sample recruitment. Y. Tsuchida contributed to critical reading and revision of the manuscript. M. Ota and K.F. designed the study and wrote the manuscript with contributions from all authors on the final version of the manuscript.

DECLARATION OF INTERESTS

N.B., N.Y., K.M, H.M., S.O., M. Okazaki, and H. Tsunoda are employees of Chugai Pharmaceutical. T.O., M. Ota, Y.N., and Y. Takeshima belong to the Social Cooperation Program, Department of Functional Genomics and Immunological Diseases, supported by Chugai Pharmaceutical. K.F. receives consulting honoraria and research support from Chugai Pharmaceutical.

Received: October 30, 2020
Revised: January 25, 2021
Accepted: March 28, 2021
Published: April 29, 2021

REFERENCES

Akizuki, S., Ishigaki, K., Kochi, Y., Law, S.M., Matsuo, K., Ohmura, K., Suzuki, A., Nakayama, M., Iizuka, Y., Koseki, H., et al. (2019). PLD4 is a genetic determinant to systemic lupus erythematosus and involved in murine autoimmune phenotypes. Ann. Rheum. Dis. *78*, 509–518.

Alasoo, K., Rodrigues, J., Mukhopadhyay, S., Knights, A.J., Mann, A.L., Kundu, K., Hale, C., Dougan, G., and Gaffney, D.J.; HIPSCI Consortium (2018). Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. Nat. Genet. *50*, 424–431.

Aletaha, D., Neogi, T., Silman, A.J., Funovits, J., Felson, D.T., Bingham, C.O., 3rd, Birnbaum, N.S., Burmester, G.R., Bykerk, V.P., Cohen, M.D., et al. (2010). 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum. *62*, 2569–2581.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Arend, W.P., Michel, B.A., Bloch, D.A., Hunder, G.G., Calabrese, L.H., Edworthy, S.M., Fauci, A.S., Leavitt, R.Y., Lie, J.T., Lightfoot, R.W., Jr., et al. (1990). The American College of Rheumatology 1990 criteria for the classification of Takayasu arteritis. Arthritis Rheum. *33*, 1129–1134.

Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R.; 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw. *67*, 1.

Battle, A., Brown, C.D., Engelhardt, B.E., and Montgomery, S.B.; GTEx Consortium; Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site—NDRI; Biospecimen Collection Source Site—RPCI; Biospecimen Core Resource—VARI; Brain Bank Repository—University of Miami Brain Endowment Bank; Leidos Biomedical—Project Management; ELSI Study; Genome Browser Data Integration &Visualization—EBI; Genome Browser Data Integration &Visualization—UCSC Genomics Institute, University of California Santa Cruz; Lead analysts; Laboratory, Data Analysis &Coordinating Center (LDACC); NIH program management; Biospecimen collection; Pathology; eQTL manuscript working group (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213.

Bohan, A., and Peter, J.B. (1975a). Polymyositis and dermatomyositis (first of two parts). N. Engl. J. Med. *292*, 344–347.

Bohan, A., and Peter, J.B. (1975b). Polymyositis and dermatomyositis (second of two parts). N. Engl. J. Med. *292*, 403–407.

Browning, B.L., Zhou, Y., and Browning, S.R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. Am. J. Hum. Genet. *103*, 338–348.

Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M.; Schizophrenia Working Group of the Psychiatric Genomics Consortium (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. *47*, 291–295.

Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. *47* (D1), D1005–D1012.

Calderon, D., Nguyen, M.L.T., Mezger, A., Kathiria, A., Müller, F., Nguyen, V., Lescano, N., Wu, B., Trombetta, J., Ribado, J.V., et al. (2019). Landscape of stimulation-responsive chromatin across diverse human immune cells. Nat. Genet. *51*, 1494–1505.

Chen, L., Ge, B., Casale, F.P., Vasquez, L., Kwan, T., Garrido-Martín, D., Watt, S., Yan, Y., Kundu, K., Ecker, S., et al. (2016). Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. Cell *167*, 1398–1414.

Cho, J.H., and Feldman, M. (2015). Heterogeneity of autoimmune diseases: pathophysiologic insights from genetics and implications for new therapies. Nat. Med. *21*, 730–738.

Chun, S., Casparino, A., Patsopoulos, N.A., Croteau-Chonka, D.C., Raby, B.A., De Jager, P.L., Sunyaev, S.R., and Cotsapas, C. (2017). Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. Nat. Genet. *49*, 600–605.

Coetzee, S.G., Coetzee, G.A., and Hazelett, D.J. (2015). motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. Bioinformatics *31*, 3847–3849.

Davenport, E.E., Amariuta, T., Gutierrez-Arcelus, M., Slowikowski, K., Westra, H.J., Luo, Y., Shen, C., Rao, D.A., Zhang, Y., Pearson, S., et al. (2018). Discovering in vivo cytokine-eQTL interactions from a lupus clinical trial. Genome Biol. *19*, 168.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Fairfax, B.P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., Jostins, L., Plant, K., Andrews, R., McGee, C., and Knight, J.C. (2014). Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science *343*, 1246949.

Farh, K.K., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh, N., Whitton, H., Ryan, R.J., Shishkin, A.A., et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature *518*, 337–343.

Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.R., Anttila, V., Xu, H., Zang, C., Farh, K., et al.; ReproGen Consortium; Schizophrenia Working Group of the Psychiatric Genomics Consortium; RACI Consortium (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. *47*, 1228–1235.

Fort, A., Panousis, N.I., Garieri, M., Antonarakis, S.E., Lappalainen, T., Dermitzakis, E.T., and Delaneau, O. (2017). MBV: a method to solve sample mislabeling and detect technical bias in large combined genotype and sequencing assay datasets. Bioinformatics *33*, 1895–1897.

Frankish, A., Diekhans, M., Ferreira, A.M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., et al. (2019). GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. *47* (D1), D766–D773.

Friedrich, M., Pohin, M., and Powrie, F. (2019). Cytokine Networks in the Pathophysiology of Inflammatory Bowel Disease. Immunity *50*, 992–1006.

Fu, B., Zhao, M., Wang, L., Patil, G., Smith, J.A., Juncadella, I.J., Zuvela-Jelaska, L., Dorf, M.E., and Li, S. (2017). RNAi Screen and Proteomics Reveal NXF1 as a Novel Regulator of IRF5 Signaling. Sci. Rep. *7*, 2683.

Fujibayashi, T., Sugai, S., Miyasaka, N., Hayashi, Y., and Tsubota, K. (2004). Revised Japanese criteria for Sjögren's syndrome (1999): availability and validity. Mod. Rheumatol. *14*, 425–434.

G'Sell, R.T., Gaffney, P.M., and Powell, D.W. (2015). A20-Binding Inhibitor of NF-κB Activation 1 is a Physiologic Inhibitor of NF-κB: A Molecular Switch for Inflammation and Autoimmunity. Arthritis Rheumatol. *67*, 2292–2302.

Gamazon, E.R., Segrè, A.V., van de Bunt, M., Wen, X., Xi, H.S., Hormozdiari, F., Ongen, H., Konkashbaev, A., Derks, E.M., Aguet, F., et al.; GTEx Consortium (2018). Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. Nat. Genet. *50*, 956–967.

Ganguly, D., Haak, S., Sisirak, V., and Reizis, B. (2013). The role of dendritic cells in autoimmunity. Nat. Rev. Immunol. *13*, 566–577.

Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. *10*, e1004383.

Godec, J., Tan, Y., Liberzon, A., Tamayo, P., Bhattacharya, S., Butte, A.J., Mesirov, J.P., and Haining, W.N. (2016). Compendium of Immune Signatures Identifies Conserved and Species-Specific Biology in Response to Inflammation. Immunity *44*, 194–206.

Griggs, R.C., Askanas, V., DiMauro, S., Engel, A., Karpati, G., Mendell, J.R., and Rowland, L.P. (1995). Inclusion body myositis and myopathies. Ann. Neurol. *38*, 705–713.

GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330.

Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. Nat. Rev. Mol. Cell Biol. *16*, 144–154.

Hochberg, M.C. (1997). Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum. *40*, 1725.

Hoogendijk, J.E., Amato, A.A., Lecky, B.R., Choy, E.H., Lundberg, I.E., Rose, M.R., Vencovsky, J., de Visser, M., and Hughes, R.A. (2004). 119th ENMC international workshop: trial design in adult idiopathic inflammatory myopathies, with the exception of inclusion body myositis, 10-12 October 2003, Naarden, The Netherlands. Neuromuscul. Disord. *14*, 337–345.

Hormozdiari, F., Kostem, E., Kang, E.Y., Pasaniuc, B., and Eskin, E. (2014). Identifying causal variants at loci with multiple signals of association. Genetics *198*, 497–508.

Hormozdiari, F., van de Bunt, M., Segrè, A.V., Li, X., Joo, J.W.J., Bilow, M., Sul, J.H., Sankararaman, S., Pasaniuc, B., and Eskin, E. (2016). Colocalization of GWAS and eQTL Signals Detects Target Genes. Am. J. Hum. Genet. *99*, 1245–1260.

Hormozdiari, F., Gazal, S., van de Geijn, B., Finucane, H.K., Ju, C.J., Loh, P.R., Schoech, A., Reshef, Y., Liu, X., O'Connor, L., et al. (2018). Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. Nat. Genet. *50*, 1041–1047.

International Study Group for Behçet's Disease (1990). Criteria for diagnosis of Behçet's disease. Lancet *335*, 1078–1080.

Ishigaki, K., Kochi, Y., Suzuki, A., Tsuchida, Y., Tsuchiya, H., Sumitomo, S., Yamaguchi, K., Nagafuchi, Y., Nakachi, S., Kato, R., et al. (2017). Polygenic burdens on cell-specific pathways underlie the risk of rheumatoid arthritis. Nat. Genet. *49*, 1120–1125.

Jenks, S.A., Cashman, K.S., Zumaquero, E., Marigorta, U.M., Patel, A.V., Wang, X., Tomar, D., Woodruff, M.C., Simon, Z., Bugrovsky, R., et al. (2018). Distinct Effector B Cells Induced by Unregulated Toll-like Receptor 7 Contribute to Pathogenic Responses in Systemic Lupus Erythematosus. Immunity *49*, 725–739.e6.

Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118–127.

Kadota, K., Ye, J., Nakai, Y., Terada, T., and Shimizu, K. (2006). ROKU: a novel method for identification of tissue-specific genes. BMC Bioinformatics *7*, 294.

Kasukawa, R. (1999). Mixed connective tissue disease. Intern. Med. *38*, 386–393.

Kim-Hellmuth, S., Aguet, F., Oliva, M., Muñoz-Aguirre, M., Kasela, S., Wucher, V., Castel, S.E., Hamel, A.R., Viñuela, A., Roberts, A.L., et al.; GTEx Consortium (2020). Cell type-specific genetic regulation of gene expression across human tissues. Science *369*, eaaz8528.

Kitching, A.R., Anders, H.J., Basu, N., Brouwer, E., Gordon, J., Jayne, D.R., Kullman, J., Lyons, P.A., Merkel, P.A., Savage, C.O.S., et al. (2020). ANCA-associated vasculitis. Nat. Rev. Dis. Primers *6*, 71.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645.

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al.; Roadmap

Epigenomics Consortium (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.

Kuznetsova, A., Brockhoff, P.B., and Christensen, R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. J. Stat. Softw. *82*, 13.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Lappalainen, T., Sammeth, M., Friedländer, M.R., 't Hoen, P.A., Monlong, J., Rivas, M.A., Gonzàlez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al.; Geuvadis Consortium (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature *501*, 506–511.

Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics *28*, 882–883.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425.

Lindley, L.E., Curtis, K.M., Sanchez-Mejias, A., Rieger, M.E., Robbins, D.J., and Briegel, K.J. (2015). The WNT-controlled transcriptional regulator LBH is required for mammary stem cell expansion and maintenance of the basal lineage. Development *142*, 893–904.

Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., et al.; International Multiple Sclerosis Genetics Consortium; International IBD Genetics Consortium (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat. Genet. *47*, 979–986.

Maecker, H.T., McCoy, J.P., and Nussenblatt, R. (2012). Standardizing immunophenotyping for the Human Immunology Project. Nat. Rev. Immunol. *12*, 191–200.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal *17*, 10–12.

Martin, F., and Chan, A.C. (2004). Pathogenic roles of B cells in human autoimmunity; insights from the clinic. Immunity *20*, 517–527.

Matsuda, S., Hammaker, D., Topolewski, K., Briegel, K.J., Boyle, D.L., Dowdy, S., Wang, W., and Firestein, G.S. (2017). Regulation of the Cell Cycle and Inflammatory Arthritis by the Transcription Cofactor *LBH* Gene. J. Immunol. *199*, 2316–2322.

Matsushita, K., Takeuchi, O., Standley, D.M., Kumagai, Y., Kawagoe, T., Miyake, T., Satoh, T., Kato, H., Tsujimura, T., Nakamura, H., and Akira, S. (2009). Zc3h12a is an RNase essential for controlling immune responses by regulating mRNA decay. Nature *458*, 1185–1190.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.

Michalska, A., Blaszczyk, K., Wesoly, J., and Bluyssen, H.A.R. (2018). A Positive Feedback Amplifier Circuit That Regulates Interferon (IFN)-Stimulated Gene Expression and Controls Type I and Type II IFN Responses. Front. Immunol. *9*, 1135.

Miyagawa-Hayashino, A., Yoshifuji, H., Kitagori, K., Ito, S., Oku, T., Hirayama, Y., Salah, A., Nakajima, T., Kiso, K., Yamada, N., et al. (2018). Increase of MZB1 in B cells in systemic lupus erythematosus: proteomic analysis of biopsied lymph nodes. Arthritis Res. Ther. *20*, 13.

Nédélec, Y., Sanz, J., Baharian, G., Szpiech, Z.A., Pacis, A., Dumaine, A., Grenier, J.C., Freiman, A., Sams, A.J., Hebert, S., et al. (2016). Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. Cell *167*, 657–669.e21.

Oestreich, K.J., and Weinmann, A.S. (2012). Master regulators or lineage-specifying? Changing views on CD4+ T cell transcription factors. Nat. Rev. Immunol. *12*, 799–804.

Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., et al.; RACI consortium; GARNET consortium (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature *506*, 376–381.

Ongen, H., Buil, A., Brown, A.A., Dermitzakis, E.T., and Delaneau, O. (2016). Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics *32*, 1479–1485.

Park, S.H., Kang, K., Giannopoulou, E., Qiao, Y., Kang, K., Kim, G., Park-Min, K.H., and Ivashkiv, L.B. (2017). Type I interferons and the cytokine TNF cooperatively reprogram the macrophage epigenome to promote inflammatory activation. Nat. Immunol. *18*, 1104–1116.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

Quach, H., Rotival, M., Pothlichet, J., Loh, Y.E., Dannemann, M., Zidane, N., Laval, G., Patin, E., Harmant, C., Lopez, M., et al. (2016). Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. Cell *167*, 643–656.e17.

Ramalingam, R., Larmonier, C.B., Thurston, R.D., Midura-Kiela, M.T., Zheng, S.G., Ghishan, F.K., and Kiela, P.R. (2012). Dendritic cell-specific disruption of TGF-β receptor II leads to altered regulatory T cell phenotype and spontaneous multiorgan autoimmunity. J. Immunol. *189*, 3878–3893.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Saha, A., and Battle, A. (2018). False positives in trans-eQTL and co-expression analyses arising from RNA-sequencing alignment errors. F1000Res. *7*, 1860.

Satoh, Y., Yokota, T., Sudo, T., Kondo, M., Lai, A., Kincade, P.W., Kouro, T., Iida, R., Kokame, K., Miyata, T., et al. (2013). The Satb1 protein directs hematopoietic stem cell differentiation toward lymphoid lineages. Immunity *38*, 1105–1115.

Schmiedel, B.J., Singh, D., Madrigal, A., Valdovino-Gonzalez, A.G., White, B.M., Zapardiel-Gonzalo, J., Ha, B., Altay, G., Greenbaum, J.A., McVicker, G., et al. (2018). Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. Cell *175*, 1701–1715.

Simeonov, D.R., Gowen, B.G., Boontanrart, M., Roth, T.L., Gagnon, J.D., Mumbach, M.R., Satpathy, A.T., Lee, Y., Bray, N.L., Chan, A.Y., et al. (2017). Discovery of stimulation-responsive immune enhancers with CRISPR activation. Nature *549*, 111–115.

Sontheimer, R.D. (2002). Would a new name hasten the acceptance of amyopathic dermatomyositis (dermatomyositis siné myositis) as a distinctive subset within the idiopathic inflammatory dermatomyopathies spectrum of clinical illness? J. Am. Acad. Dermatol. *46*, 626–636.

Sontheimer, R.D. (2017). MDA5 autoantibody-another indicator of clinical diversity in dermatomyositis. Ann. Transl. Med. *5*, 160.

Stegle, O., Parts, L., Piipari, M., Winn, J., and Durbin, R. (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat. Protoc. *7*, 500–507.

Sul, J.H., Han, B., Ye, C., Choi, T., and Eskin, E. (2013). Effectively identifying eQTLs from multiple tissues by combining mixed model and meta-analytic approaches. PLoS Genet. *9*, e1003491.

Sun, J., Nishiyama, T., Shimizu, K., and Kadota, K. (2013). TCC: an R package for comparing tag count data with robust normalization strategies. BMC Bioinformatics *14*, 219.

Tadaka, S., Saigusa, D., Motoike, I.N., Inoue, J., Aoki, Y., Shirota, M., Koshiba, S., Yamamoto, M., and Kinoshita, K. (2018). jMorp: Japanese Multi Omics Reference Panel. Nucleic Acids Res. *46* (D1), D551–D557.

Tawfik, D.M., Lankelma, J.M., Vachot, L., Cerrato, E., Pachot, A., Wiersinga, W.J., and Textoris, J. (2020). Comparison of host immune responses to LPS

in human using an immune profiling panel, in vivo endotoxemia versus ex vivo stimulation. Sci. Rep. *10*, 9918.

Trynka, G., Sandor, C., Han, B., Xu, H., Stranger, B.E., Liu, X.S., and Raychaudhuri, S. (2013). Chromatin marks identify critical cell types for fine mapping complex trait variants. Nat. Genet. *45*, 124–130.

Tsuchiya, H., Ota, M., Sumitomo, S., Ishigaki, K., Suzuki, A., Sakata, T., Tsuchida, Y., Inui, H., Hirose, J., Kochi, Y., et al. (2020). Parsing multiomics landscape of activated synovial fibroblasts highlights drug targets linked to genetic risk of rheumatoid arthritis. Ann. Rheum. Dis., annrheumdis-2020-218189.

Urbut, S.M., Wang, G., Carbonetto, P., and Stephens, M. (2019). Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. Nat. Genet. *51*, 187–195.

van Arensbergen, J., Pagie, L., FitzPatrick, V.D., de Haas, M., Baltissen, M.P., Comoglio, F., van der Weide, R.H., Teunissen, H., Võsa, U., Franke, L., et al. (2019). High-throughput identification of human SNPs affecting regulatory element activity. Nat. Genet. *51*, 1160–1169.

van den Hoogen, F., Khanna, D., Fransen, J., Johnson, S.R., Baron, M., Tyndall, A., Matucci-Cerinic, M., Naden, R.P., Medsger, T.A., Jr., Carreira, P.E., et al. (2013). 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism collaborative initiative. Arthritis Rheum. *65*, 2737–2747.

van der Wijst, M.G.P., Brugge, H., de Vries, D.H., Deelen, P., Swertz, M.A., and Franke, L.; LifeLines Cohort Study; BIOS Consortium (2018). Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. Nat. Genet. *50*, 493–497.

van Kempen, T.S., Wenink, M.H., Leijten, E.F., Radstake, T.R., and Boes, M. (2015). Perception of self: distinguishing autoimmunity from autoinflammation. Nat. Rev. Rheumatol. *11*, 483–492.

Võsa, U., Claringbould, A., Westra, H.-J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Kasela, S., et al. (2018). Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. bioRxiv. https://doi.org/10.1101/447367.

Wahren-Herlenius, M., and Dörner, T. (2013). Immunopathogenic mechanisms of systemic autoimmune disease. Lancet *382*, 819–831.

Watts, R., Lane, S., Hanslik, T., Hauser, T., Hellmich, B., Koldingsnes, W., Mahr, A., Segelmark, M., Cohen-Tervaert, J.W., and Scott, D. (2007). Development and validation of a consensus methodology for the classification of the ANCA-associated vasculitides and polyarteritis nodosa for epidemiological studies. Ann. Rheum. Dis. *66*, 222–227.

Whitfield, M.L., George, L.K., Grant, G.D., and Perou, C.M. (2006). Common markers of proliferation. Nat. Rev. Cancer *6*, 99–106.

Yamaguchi, M., Ohta, A., Tsunematsu, T., Kasukawa, R., Mizushima, Y., Kashiwagi, H., Kashiwazaki, S., Tanimoto, K., Matsumoto, Y., Ota, T., et al. (1992). Preliminary criteria for classification of adult Still's disease. J. Rheumatol. *19*, 424–430.

Yee, C.S., Farewell, V.T., Isenberg, D.A., Griffiths, B., Teh, L.S., Bruce, I.N., Ahmad, Y., Rahman, A., Prabu, A., Akil, M., et al. (2011). The use of Systemic Lupus Erythematosus Disease Activity Index-2000 to define active disease and minimal clinically meaningful change based on data from a large cohort of systemic lupus erythematosus patients. Rheumatology (Oxford) *50*, 982–988.

Zhang, H., Zhao, C., Wang, S., Huang, Y., Wang, H., Zhao, J., and Yang, N. (2015). Anti-dsDNA antibodies induce inflammation via endoplasmic reticulum stress in human mesangial cells. J. Transl. Med. *13*, 178.

Zhernakova, D.V., Deelen, P., Vermaat, M., van Iterson, M., van Galen, M., Arindrarto, W., van 't Hof, P., Mei, H., van Dijk, F., Westra, H.J., et al. (2017). Identification of context-dependent expression quantitative trait loci in whole blood. Nat. Genet. *49*, 139–145.

Please cite this article in press as: Ota et al., Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases, Cell (2021), https://doi.org/10.1016/j.cell.2021.03.056

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
See Table S9 for flow-cytometry antibodies	N/A	N/A
Critical commercial assays		
MACSxpress Neutrophil Isolation Kit, human	Miltenyi Biotec	Cat#130-104-434
EasySep Direct Human Neutrophil Isolation Kit	STEMCELL Technologies	Cat#19666
RNeasy Micro Kit	QIAGEN	Cat#74004
MagMAX-96 Total RNA Isolation Kit	Thermo Fisher Scientific	Cat#AM1830
RNeasy Mini Kits	QIAGEN	Cat#74104
SMART-seq v4 Ultra Low Input RNA Kit for Sequencing	Takara Bio	Cat#634894
QIAmp DNA Blood Midi Kit	QIAGEN	Cat#51185
TruSeq DNA PCR-Free Library prep kit	Illumina	Cat#20015963
HiSeq SBS Kit v4	Illumina	Cat#FC-401-4003
Infinium OmniExpressExome-8	Illumina	Cat#20024679
Deposited data		
Web-based data browser	This paper	https://www.immunexut.org/
Gene expression data	This paper	National Bioscience Database Center (NBDC): E-GEAD-397
Expression QTL summary statistics (conditional eQTL analysis)	This paper	NBDC: E-GEAD-398
Expression QTL summary statistics (nominal eQTL statistics, including non-significant associations)	This paper	NBDC: E-GEAD-420
Human reference genome NCBI build 38, GRCh38	Genome Reference Consortium	https://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/
1000 Genomes Project reference panel (phase3)	Auton et al., 2015	https://www.internationalgenome.org/data
GENCODE annotation v27 (GRCh38)	Frankish et al., 2019	https://www.gencodegenes.org/
NIH Roadmap Epigenomics Mapping Consortium	Kundaje et al., 2015	https://egg2.wustl.edu/roadmap/web_portal/processed_data.html
Immune cell ATAC-seq count data	Calderon et al., 2019	GEO: GSE118189
Immune cell ATAC-seq bigwig files	Calderon et al., 2019	https://s3.amazonaws.com/muellerf/data/trackhubs/immune_atlas/hub.txt
Stimulated fibroblast-like synoviocytes gene expression data	Tsuchiya et al., 2020	NBDC: hum0207.v1.RNA.v1
GTEx v8 eQTL summary statistics	GTEx consortium, 2020	https://gtexportal.org/home/datasets
eQTL summary statistics of bulk immune-cell eQTL	Ishigaki et al., 2017	NBDC: hum0099.v1.eqtl.v1
eQTL summary statistics of DICE database	Schmiedel et al., 2018	https://dice-database.org/downloads
Regulatory variants in Survey of regulatory elements (SuRE) reporter technology	van Arensbergen et al., 2019	https://osf.io/w5bzq/wiki/home/?view
NHGRI-EBI GWAS Catalog	Buniello et al., 2019	https://www.ebi.ac.uk/gwas/
eQTL summary statistics of eQTLGen consortium	Võsa et al., 2018	https://www.eqtlgen.org/
Tohoku Medical Megabank Organization	Tadaka et al., 2018	https://jmorp.megabank.tohoku.ac.jp/202008/
Cross-mappability and gene-mappability	Saha and Battle, 2018	https://figshare.com/collections/Mappability_Resources/4297352/4
GWAS summary statistics for LDSC	Finucane et al., 2015	https://data.broadinstitute.org/alkesgroup/sumstats_formatted/
MSigDB hallmark gene set collection	Liberzon et al., 2015	https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp

(*Continued on next page*)

Please cite this article in press as: Ota et al., Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases, Cell (2021), https://doi.org/10.1016/j.cell.2021.03.056

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
PLINK (v1.9)	Purcell et al., 2007	https://www.cog-genomics.org/plink2
cutadapt (v1.16)	Martin, 2011	https://cutadapt.readthedocs.io/en/stable/index.html
fastx-toolkit (v0.0.14)	N/A	http://hannonlab.cshl.edu/fastx_toolkit/
STAR (v2.5.3)	Dobin et al., 2013	https://github.com/alexdobin/STAR
HTSeq (v 0.11.2.)	Anders et al., 2015	https://github.com/htseq/htseq
bcl2fastq (v 2.20.0)	Illumina	https://jp.support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html
GATK (v 4.0.6.0)	DePristo et al., 2011	https://gatk.broadinstitute.org/hc/en-us
BWA-MEM (v 0.7.17)	Li and Durbin, 2009	http://bio-bwa.sourceforge.net/
BEAGLE (v 5.1)	Browning et al., 2018	https://faculty.washington.edu/browning/beagle/beagle.html#citation
QTLtools (v1.2)	Fort et al., 2017	https://qtltools.github.io/qtltools/
CAVIAR (v2.0.0)	Hormozdiari et al., 2014	https://github.com/fhormoz/caviar
eCAVIAR (v2.0.0)	Hormozdiari et al., 2016	https://github.com/fhormoz/caviar
CIRCOS (v0.69-9)	Krzywinski et al., 2009	http://circos.ca/
LDSC (v1.0.1)	Bulik-Sullivan et al., 2015	https://github.com/bulik/ldsc
Meta-Tissue	Sul et al., 2013	http://genetics.cs.ucla.edu/metatissue/
R (v3.4 or v3.5, depending on the time of the analysis)	N/A	https://www.r-project.org/
edgeR (v3.24.3)	Robinson et al., 2010	https://bioconductor.org/packages/release/bioc/html/edgeR.html
sva (v3.30.1)	Leek et al., 2012	https://www.bioconductor.org/packages/release/bioc/html/sva.html
lme4 (v1.1.21)	Bates et al., 2015	https://cran.r-project.org/web/packages/lme4/index.html
TCC (v1.22.1)	Sun et al., 2013	https://www.bioconductor.org/packages/release/bioc/html/TCC.html
WGCNA (v1.67)	Langfelder and Horvath, 2008	https://cran.r-project.org/web/packages/WGCNA/index.html
limma(v3.38.3)	Ritchie et al., 2015	http://bioconductor.org/packages/release/bioc/html/limma.html
mashr (v0.2.21)	Urbut et al., 2019	https://github.com/stephenslab/mashr
peer (v1.0)	Stegle et al., 2012	https://github.com/PMBio/peer
lmerTest (v3.1.1)	Kuznetsova et al., 2017	https://cran.r-project.org/web/packages/lmerTest/index.html
motifbreakR (v1.12.0)	Coetzee et al., 2015	https://bioconductor.org/packages/release/bioc/html/motifbreakR.html
coloc (v3.2.1)	Giambartolomei et al., 2014	https://cran.r-project.org/web/packages/coloc/index.html
Other		
Cell-type-specific expressed genes	This paper	Table S2
Genes especially dysregulated in specific diseases	This paper	Table S3
Context-dependent eQTLs	This paper	Table S4
Expression QTL effects of Immune disease GWAS variants	This paper	Table S5

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Keishi Fujio (fujiok-int@h.u-tokyo.ac.jp).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Gene expression data and the summary statistics of eQTL analysis were deposited in the National Bioscience Database Center (NBDC) Human Database (https://humandbs.biosciencedbc.jp/en/) with the accession number of E-GEAD-397, E-GEAD-398 and E-GEAD-420. These data can be downloaded without restriction. We used publicly available software for the analyses. Custom code is available from the corresponding authors upon reasonable request. Data can also be browsed at our website at https://www.immunexut.org/.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

This study was approved by the Ethics Committees of the University of Tokyo, and written informed consent was obtained from all subjects enrolled. Age and sex of all the participants are deposited in NBDC with the same accession number as denoted above. Inclusion and exclusion criteria for our cohort is described in Method Details.

METHOD DETAILS

Inclusion and exclusion criteria

ImmuNexUT consists of patients with 10 categories of IMDs and healthy volunteers.

The systemic sclerosis (SSc) cohort comprised of patients who had attended the Department of Allergy and Rheumatology at University of Tokyo Hospital or Department of Rheumatology at Tokyo Metropolitan Komagome Hospital, who met the 2013 American College of Rheumatology (ACR)/ European League Against Rheumatism (EULAR) classification criteria for Systemic Sclerosis (van den Hoogen et al., 2013) and who were taking no more than 20 mg prednisolone daily.

The systemic lupus erythematosus (SLE) cohort comprised of patients who had attended the Department of Allergy and Rheumatology at University of Tokyo Hospital, Division of Rheumatic Diseases at National Center for Global Health and Medicine or Immuno-Rheumatology Center at St. Luke's International Hospital, who met the 1997 revised version of ACR SLE criteria (Hochberg, 1997). Of the enrolled patients, 63% had active disease (Yee et al., 2011) with SLEDAI-2000 > 3, and the other cases had inactive disease at the time of blood withdrawal.

The idiopathic inflammatory myopathy (IIM) cohort comprised of patients who had attended to the Department of Allergy and Rheumatology at University of Tokyo Hospital or Division of Rheumatology at the Jikei University Hospital and who met either one of following criteria: Bohan and Peter criteria (Bohan and Peter, 1975a, 1975b), the European Neuromuscular Center criteria (Hoogendijk et al., 2004), Sontheimer criteria (Sontheimer, 2002) or Griggs criteria (Griggs et al., 1995). Patients taking more than 11 mg prednisolone daily were excluded. Clinical diagnosis of the patients was dermatomyositis in 30, clinically amyopathic dermatomyositis in 15, polymyositis in 11, immune-mediated necrotizing myopathy in 4 and inclusion body myositis in 2, and the other 3 cases were not clearly classified into these categories. 23 cases had active disease and required subsequent initiation or increase of immunomodulatory drugs.

Behçet's disease (BD) patients were recruited from the Department of Allergy and Rheumatology at University of Tokyo Hospital. Patients who met both the criteria of the Behçet's Disease Research Committee of Japan and the International Criteria for Behçet's Disease (International Study Group for Behçet's Disease, 1990) were included. None of the participants were receiving biologics, and none of the participants was in the acute or active phase.

Sjögren's syndrome (SjS) patients were recruited from the Department of Allergy and Rheumatology at University of Tokyo Hospital. Patients who met the criteria of the Sjögren's syndrome Research Committee of Japan (Fujibayashi et al., 2004) and clinically diagnosed as primary Sjögren's syndrome were included.

Rheumatoid arthritis (RA) patients were recruited from the Department of Allergy and Rheumatology at University of Tokyo Hospital. Patients who met the 2010 ACR/EULAR classification criteria for rheumatoid arthritis (Aletaha et al., 2010) were included. DAS28ESR (median, interquartile range) was 4.99 (4.35-5.97), indicating that the majority of the patients had moderate to high disease activity.

Adult-onset Still's disease (AOSD) patients were recruited from the Department of Allergy and Rheumatology at University of Tokyo Hospital. Patients who met the Yamaguchi's criteria (Yamaguchi et al., 1992) and who were taking no more than 10 mg prednisolone daily were enrolled.

ANCA-associated vasculitis (AAV) patients were recruited from the Department of Allergy and Rheumatology at University of Tokyo Hospital and Immuno-Rheumatology Center at St. Luke's International Hospital. Patients classified as granulomatosis with polyangiitis or microscopic polyangiitis following Watts algorithm (Watts et al., 2007) and positive for either myeloperoxidase-anti-neutrophil cytoplasmic antibody or proteinase3-anti-neutrophil cytoplasmic antibody were included. Patients taking more than 10 mg prednisolone daily were excluded.

Please cite this article in press as: Ota et al., Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases, Cell (2021), https://doi.org/10.1016/j.cell.2021.03.056

Patients with Takayasu arteritis (TAK) were recruited from the Department of Allergy and Rheumatology at University of Tokyo Hospital. Patients who fulfilled the American College of Rheumatology 1990 criteria for the classification of Takayasu arteritis (Arend et al., 1990) were included. Patients clinically diagnosed as giant cell arteritis or receiving biologics were excluded.

Patients with mixed connective tissue disease (MCTD) were recruited from the Department of Allergy and Rheumatology at University of Tokyo Hospital. Patients who fulfilled the 1996 revised version of Kasukawa's criteria (Kasukawa, 1999) were included. Patients clinically diagnosed as overlap syndrome were excluded.

Patients with apparent malignancies or active infections at the time of enrollment were excluded.

The inclusion criteria for healthy volunteers were people with no apparent co-morbidities, no direct family history of autoimmune diseases, and no use of prescription drugs or supplements. Age and sex were matched with patient cohort as much as possible.

Samples were collected from the University of Tokyo, the Jikei University School of Medicine, St. Luke's International Hospital, National Center for Global Health, and Medicine and Tokyo Metropolitan Komagome Hospital with approval by the Ethics Committees at each site.

Sample collection and processing

We modified our sample collection protocol during this study. Samples before and after modification are termed phase 1 and phase 2 samples, respectively. In both phases, peripheral blood mononuclear cells (PBMCs) were isolated by density gradient separation with Ficoll-Paque (GE healthcare) immediately after blood drawing. Erythrocytes were lysed with potassium ammonium chloride buffer, and non-specific binding was blocked with Fc-gamma receptor antibodies. In phase 1, we sorted PBMCs into 19 immune cell subsets (Table S8) with purity > 99% using a MoFlo XDP instrument (Beckman Coulter). Sorted cells were lysed and stored at −80°C. Total RNA was extracted using RNeasy Micro Kits (QIAGEN). Libraries for RNA-seq were prepared using SMART-seq v4 Ultra Low Input RNA Kit (Takara Bio). In phase 2, we sorted PBMCs into 26 immune cell subsets (Table S8) with purity > 99% using a 14-color cell sorter BD FACSAria Fusion (BD Biosciences). Sorted cells were lysed and stored at −80°C. Total RNA was extracted using MagMAX-96 Total RNA Isolation Kits (Thermo Fisher Scientific). Libraries for RNA-seq were prepared with the same procedures used in phase 1. In both phases, we intended to collect 5,000 cells with at least 1,000 cells per subset (5,000 cells were collected for > 80% of samples). We followed previously reported immune cell definitions provided by the Human Immunology Project (Maecker et al., 2012) for the flowcytometry staining panel with slight modification due to the availability of labeled antibodies in each phase (Tables S8 and S9). In addition, neutrophils were collected with EasySep Direct Human Neutrophil Isolation Kits (STEMCELL Technologies) or MACSxpress Neutrophil Isolation Kits human (Miltenyi Biotec) with an aim of 2×10^6 cells, lysed and stored at −80°C, followed by RNA isolation with an RNeasy Mini Kits (QIAGEN) and library preparation with SMART-seq v4 Ultra Low Input RNA Kits (Takara Bio).

Genomic DNA was isolated from peripheral blood using QIAmp DNA Blood Midi kit (QIAGEN) and libraries for whole genome sequencing were prepared using TruSeq DNA PCR-Free Library prep kit (Illumina).

QUANTIFICATION AND STATISTICAL ANALYSIS

RNA sequencing

Samples failing any of the quality control steps (RNA quality and quantity, PCR amplification, library fragment size) were eliminated from further downstream steps. In total, we performed RNA-seq for 10,102 samples (2,942 samples in phase 1 and 7,160 samples in phase 2) from 423 donors. Libraries were sequenced on the HiSeq 2500 Illumina platform to obtain 100-bp paired-end reads with HiSeq SBS Kit v4 (Illumina), generating a total of over 211 billion reads.

QC of gene expression data

From sequenced reads, adaptor sequences were trimmed using cutadapt (v1.16). In addition, 3′- ends with low-quality bases (Phred quality score < 20) were trimmed using the fastx-toolkit (v0.0.14). Reads containing more than 20% low-quality bases were removed. Subsequently, reads were aligned against the GRCh38 reference sequence using STAR (v2.5.3) (Dobin et al., 2013) in two-pass mode with Gencode version 27 annotations (Frankish et al., 2019). For gene level quantification, we combined all isoforms of a gene into a single transcript as described elsewhere (Battle et al., 2017) and used it for annotations. We excluded samples with uniquely mapped read rates < 90% (with the exception of < 70% for plasmablasts and < 85% for the other B cell subsets, which have a nonnegligible number of unmapped reads because of highly variable B cell receptors) or unique read counts < 6×10^6. A median of 10.1 million paired fragments were uniquely mapped per sample, and only concordantly and uniquely mapped read pairs were used for analysis. Expression was quantified using HTSeq (v 0.11.2.) (Anders et al., 2015).

For QC of the expression data, in each cell population, we filtered low count genes (< 10 in > 90% of samples), normalized between samples with a trimmed mean of M values (TMM) implemented in edgeR (Robinson et al., 2010) software, converted to log-transformed count per million (CPM), removed batch effects using ComBat software (Johnson et al., 2007) and computed inter-sample Spearman's correlations of expression levels between each sample and the remaining samples from the same cell subset. We excluded 72 samples with mean correlation coefficients less than 0.9. Moreover, we excluded all the samples from 5 donors from whom 3 or more samples were excluded as outliers.

Please cite this article in press as: Ota et al., Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases, Cell (2021), https://doi.org/10.1016/j.cell.2021.03.056

Whole genome sequencing and variant calling

Whole genomes were sequenced on Illumina's HiSeq X with 151-bp pair-end reads. The mean coverage was 41x. The BCL binary format was converted into the FASTQ format using bcl2fastq (v 2.20.0) (Illumina). Data processing was performed based on the standardized best-practice method proposed by GATK (v 4.0.6.0) (DePristo et al., 2011). Briefly, FASTQ files were converted to SAM files with FastqToSam, adapters marked with MarkIlluminaAdapters, and reads were aligned to the reference human genome (GRCh38) using BWA-MEM (v 0.7.17, default parameters) (Li and Durbin, 2009). The duplicate reads were marked with MarkDuplicates. Base quality score recalibration was performed with default references. Variants were called with HaplotypeCaller with "–pcr-indel-model NONE" parameter. Called GVCFs from each sample were jointly genotyped and merged using GenomicsDBImport, GenotypeGVCFs and MergeVcfs. Variants with ExcessHet > 54.69 were filtered. Variant quality score recalibration was performed for both single nucleotide variants (SNVs) and indels with GATK recommended references. After applying variant quality score recalibration to merged VCF, genotype quality (GQ) scores were recalculated for all calls based on allele frequencies in the 1000 Genomes Project, phase 3, East Asian (EAS) population using CalculateGenotypePosteriors. Samples with genotyping call rates < 99% were removed. Multi-allelic sites were split into bi-allelic. Calls with GQ < 20 and/or DP < 5 were assigned to missing. Variants with genotyping call rate < 85% and/or HWE P value < 1.0×10^{-6} were removed. Finally, we used BEAGLE (v 5.1) (Browning et al., 2018) to impute missing genotypes. Combined with the QC of RNA-seq data, 416 samples were utilized for the subsequent analysis. Variants with minor allele frequency < 1% were excluded, yielding a total of 8,105,611 autosomal variants (7,154,278 SNVs and 951,333 indels).

Validation of WGS calls

For 2 donors, we obtained genome-wide SNP genotypes by using the SNP microarray approach (Infinium OmniExpressExome-8 [Illumina]) and confirmed high concordance (> 99.98%) with WGS calls.

We performed principal component analysis (PCA) of called genotype data jointly with 1000G phase 3 samples after removing the HLA region and pruning. As expected, our Japanese samples clustered together with 1000G EAS samples (Figure S3A).

We compared the allele frequencies of our genotype data with Tohoku Medical Megabank Organization (ToMMo) data (Tadaka et al., 2018), a large scale WGS reference panel from Japan, and there was a good concordance in allele frequencies for SNVs (Pearson's correlation 0.998, Figure S3B) and indels (Pearson's correlation 0.995, Figure S3C).

Analyzed samples

For subsequent analyses, we utilized 9,852 RNA-seq samples from 416 donors that had passed RNA-seq QC and had corresponding WGS data that had also passed QC.

Variance decomposition

For those genes whose expression was more than 5 in at least 80% of samples in at least one cell type, the count data from all samples (except for neutrophils, which were collected without using FACS) were normalized between samples with TMM, converted to CPM, log transformed and batch effects removed (i.e., study phases) with Combat (Leek et al., 2012) software. Normalized data were fit to the following linear mixed model using the lme4 package (Bates et al., 2015):

$$Expression \approx (1|Celltype) + (1|Diagnosis) + (1|Individual)$$

Subsequently, we calculated the variance explained by each variable for each gene. Standard deviation was estimated by bootstrapping of samples 1,000 times with the bootMer function.

Hierarchical clustering

Twenty-five immune cell subsets that did not show population overlap with other sorted immune cells (i.e., all subsets except for Mem CD4, Mem CD8 and CL Mono) were utilized for hierarchical clustering. In each cell subset, we filtered genes expressed at low levels, and counts were normalized, and cleaned for batch effects with Combat software. Then, Spearman's correlation distance of the top 5,000 variable genes was used for hierarchical clustering with Ward's method.

Cell type-specific genes

Cell type-specific genes were identified using 25 immune cell subsets denoted above from healthy volunteers in phase 2. We filtered genes with low expression levels, and counts were normalized, averaged and cell type-specific genes were identified based on their Shannon entropies using the ROKU (Kadota et al., 2006) function in TCC package (Sun et al., 2013).

Transcriptome signature construction

Nineteen cell subsets were consistently collected from almost all the individuals (> 395 individuals; all subsets except for LDG, Fr. I nTreg, Fr. III T, CM CD8, TEMRA CD8, EM CD8, Mem CD8, Int Mono and NC Mono). Those subsets were utilized for transcriptome signature comparisons. In each immune cell data, genes expressed at low levels were filtered and residual expression data were normalized with TMM, converted to CPM, log transformed and batch effects removed with Combat software.

Please cite this article in press as: Ota et al., Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases, Cell (2021), https://doi.org/10.1016/j.cell.2021.03.056

We applied the weighted gene co-expression network analysis (WGCNA) (Langfelder and Horvath, 2008) algorithm to normalized data of each cell type with a "signed network" option and soft threshold power with the adjacency matrix set to 8. After making modules, gene modules whose first eigenvectors were highly correlated (Pearson's correlation > 0.9) were merged and modules with more than 2,000 genes were removed in each subset to reduce the redundancy and uncertainty of the modules. This procedure generated a total of 391 gene modules (median of 22 for each cell type) that were used for subsequent analyses.

Eigenvectors of each module were then fit to the following model with linear regression:

$$Eigenvector \approx Diagnosis + \log(PSLmg + 0.1) + \log(Age) + StudyPhase + Sex$$

None of the modules showed a significant association with the study phase, meaning successful cleaning of batch effects by the previous procedures. We set healthy volunteers as the control for the "Diagnosis" term. After multiple test corrections, 238 out of 391 modules were significantly associated with at least 1 diagnosis (FDR < 0.01), whose standardized regression coefficients were utilized for clustering analysis (Figure 1E). Clinical diagnosis was hierarchically clustered based on Pearson's correlation distance of standardized regression coefficients, while modules were grouped by k-means clustering.

For clustering analysis of individuals (Figure S2B), we utilized 45 modules which were significantly upregulated in some IMDs and annotated to specific pathways by cytokine pathway enrichment analysis. Eigenvectors of modules were used for k-means clustering.

Cytokine pathway enrichment analysis

To associate gene modules to biological pathways, we utilized an independent dataset which we reported recently (Tsuchiya et al., 2020). In this dataset, fibroblast-like synoviocytes (FLS) from RA and osteoarthritis patients were treated with 8 kinds of cytokines *in vitro*, and RNA-seq was performed on unstimulated or stimulated samples. In order to make gene modules which are associated with specific stimulation patterns, we performed WGCNA with the FLS dataset with the same parameters as denoted above except for setting a soft threshold power to 15, and eigenvectors were fit to the following model with linear regression:

$$Eigenvector \approx Stimulation + Diagnosis$$

We set non-stimulated samples as the controls for the "Stimulation" term, and standardized regression coefficients of each stimulation were compared for each module (Figure S2A), which showed stimulation-dependent expression perturbation patterns.

We utilized FLS gene modules that were significantly associated with at least 1 stimulation condition (FDR < 0.01) for comparison with ImmuNexUT gene modules. Gene set enrichment was evaluated with a one-sided Fisher's exact test.

Genes especially dysregulated in specific diseases

We sought to identify genes which are especially up- or downregulated in specific IMDs. First, we identified differentially expressed genes (DEGs) by comparing each IMD versus healthy volunteers in 19 cell subsets which were collected from almost all the individuals. The limma package (Ritchie et al., 2015) with voom transformation was utilized. Sex, age, prednisolone dosage and study phase were treated as covariates. To improve the power to estimate disease specific and shared DEGs, we utilized mash (Urbut et al., 2019) software with effect sizes of each comparison and their standard deviations as input. Genes with FDR < 0.05 in any of the disease-by-disease comparisons were analyzed. All the other genes were utilized to fit the mash model. Mash outputs improved effect estimates of each comparison and the local false sign rate (LFSR), which is analogous to FDR while considering the sign of the association. We made a list of genes especially dysregulated in each disease by applying the following 3 criteria; i) LFSR was < 0.05, ii) diseases with LFSR < 0.05 were 2 or less, iii) the absolute values of effect estimates are 1.4 times or more higher than those of the other diseases.

Data normalization and eQTL analysis

After sample QC, genes expressed at low levels (< 5 count in more than 80% samples or < 0.5 CPM in more than 80% samples) were filtered out in each cell subset. The residual expression data were normalized between samples with TMM, converted to CPM and then normalized across samples using an inverse normal transform. A Probabilistic Estimation of Expression Residuals (PEER) method (Stegle et al., 2012) was applied to normalized expression data to infer hidden covariates. We empirically tested for the number of eGenes with incrementation of PEER factors. To maximize the number of eGenes and avoid potential overfitting, 50 PEER factors for cell subsets with > 350 donors and 30 PEER factors for cell subsets with $\leq$ 350 donors were chosen for correction of hidden covariates (Figures S3D and S3E). The top 2 genetic principal components, sample collection phase, clinical diagnosis, sex and PEER factors were utilized as covariates for eQTL analysis. Mem CD8s, which were collected in phase 1 and divided into CM CD8 and EM CD8 in phase 2, were analyzed jointly with EM CD8 for eQTL analysis because the majority of the Mem CD8 population consisted of EM CD8.

To evaluate the concordance of WGS and RNA-seq data, we used the mbv module from QTLtools (Fort et al., 2017) and confirmed a perfect match of RNA-seq and WGS labels.

Please cite this article in press as: Ota et al., Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases, Cell (2021), https://doi.org/10.1016/j.cell.2021.03.056

For each cell subset, we used a QTLtools permutation pass with 10,000 permutations to obtain gene-level nominal *P value* thresholds corresponding to FDR < 0.05. We subsequently performed forward-backward stepwise regression eQTL analysis with a QTLtools conditional pass.

Epigenome mark enrichment analysis

To evaluate the enrichment of top eVariants to epigenome marks (differentially accessible peaks or Roadmap annotation), we constructed random variants sets that were matched to top eVariants based on distance to the nearest TSS, minor allele frequency and chromosome. We compared the overlap with epigenome marks between random variants sets and real ones.

The chromatin state data were obtained from the Roadmap Epigenome Project. We utilized 18-state models. We considered TssA, TssAFlnk, TssFlnk, TssFlnkD and TssFlnkU as promoter regions, and Enh, EnhA1, EnhA2, EnhG, EnhG1, EnhG2 and EnhWk as enhancer regions. For analysis of enhancer or promoter enrichment (Figure 2B), we utilized 17 immune cell subsets that had relevant cell annotations in Roadmap project data (i.e., CL Mono [Roadmap ID: E029], CD16p Mono [E029], DN B[E032], Fr. II eTreg [E044], Mem CD4 [E037], Mem CD8 [E048], Naive B [E032], Naive CD4 [E039], Naive CD8 [E047], NK [E046], Plasmablast [E032], SM B [E032], Tfh [E045], Th1 [E045], Th17 [E042], Th2 [E045] and USM B [E032]).

ATAC-seq data of immune cells (Calderon et al., 2019) were obtained from GEO under accession GSE118189. For comparison with each cell eQTL data, we utilized read count data of non-stimulated immune cell subsets which had analogous cell types in ImmuNexUT. For each immune cell type, the read count of peaks with mean CPM > 1 were compared with the other cell types using edgeR, and the top 14,000 differentially accessible peaks were utilized for enrichment analysis. For comparison with IMD-specific eQTLs, we utilized differentially accessible peaks after stimulation (Calderon et al., 2019). Calderon et al., strongly stimulated various kinds of immune cells *in vitro* and cataloged differentially accessible peaks after stimulation. We aggregated differential peaks induced after stimulation (FDR < 0.01 and log Fold change > 0) in any of the cell types and utilized them as "peaks induced by stimulation." For comparison, we utilized peaks in non-stimulated immune cells that were not significantly changed after stimulations as "unchanged peaks." For visualization of ATAC-seq peaks, we utilized bigwig files provided by the authors (https://s3.amazonaws.com/muellerf/data/trackhubs/immune_atlas/hub.txt) and visualized them with the UCSC genome browser (http://genome.ucsc.edu/.).

Comparison of eQTL effect sizes

In order to compare effect sizes and cell type specificities of eQTLs, we utilized mash (Urbut et al., 2019) for top eVariants. The analysis was performed with eQTL beta coefficients and their standard errors as inputs. Randomly selected SNP-gene pairs (220,000) were used to fit the mash model. If the gene was not expressed in the subset, the beta coefficient was set to 0. Effect size estimates and LFSR outputted by mash were used as metrics of QTL magnitude and significance.

Comparison of eQTL effects between IMD patients and healthy volunteers

For comparison of eQTL effects between IMD patients and healthy volunteers, we performed forward-backward stepwise eQTL analysis in healthy volunteer group and IMD patient group separately. For this comparison, cell types were limited to 20 cell types that were obtained from almost all the donors from both groups, variants with minor allele frequency > 0.1 in both groups, and genes with $\geq$ 5 count in more than 20% samples besides $\geq$ 0.5 CPM in more than 20% samples in both groups. Obtained eQTL beta coefficients and their standard errors for top eVariants in either group were utilized as inputs to mash (in total 40 cell type-group pairs). For stringency, we defined eQTLs with LFSR < 5% in IMD patients and > 25% in healthy volunteers as IMD specific eQTLs, and vice versa. Expression QTLs with LFSR < 5% in both groups were defined as eQTLs significant in both. For the comparative analysis (Figure 2F), the effect estimates of top eVariants of each cell type were compared between IMD patients and healthy volunteers.

Comparison with other data

We compared beta coefficients of our eQTLs with the eQTLs in a study of 105 healthy Japanese volunteers (Ishigaki et al., 2017). Naive CD4, naive B, NK and CL monocyte eQTLs were compared with bulk CD4, B, NK and monocyte eQTLs from the previous study. Only top eVariants in this study that achieved FDR < 0.05 in the previous study were considered (Figure 3A).

In addition, the latest dataset of whole blood cis-eQTLs consisting of > 30,000 Europeans (Võsa et al., 2018) was downloaded from the eQTLGen website https://www.eqtlgen.org/ on 1/29/2020 and compared with our dataset. Only genes and variants analyzed in both datasets were utilized for all the comparison analyses.

Fine-mapping of eQTLs

We fine-mapped eQTL variants of GTEx v8 whole blood data of European ancestry and ImmuNexUT classical monocytes and neutrophils. In order to assess loci whose eQTL causal variants were shared between 2 datasets, we limited our analysis to eVariants that were polymorphic in both, eGenes that were significant in both and the eGene whose top eVariant in GTEx had a P value < 1×10^{-5} in ImmuNexUT. These limitations resulted in 2,575 genes for classical monocytes and 2,359 genes in neutrophils. For fine-mapping of GTEx and our eQTLs, we ranked eVariants by *P value*s in GTEx and analyzed the top 50 variants with CAVIAR (Hormozdiari et al., 2014) while permitting up to 5 causal variants. For joint fine-mapping analysis, we utilized eCAVIAR (Hormozdiari et al., 2016) for the same set of variants while permitting up to 5 causal variants. Our whole genome data were utilized for estimating LD for Japanese,

and the 1000 genomes project phase 3 European samples were used for estimating LD for Europeans. We compared the number of variants in fine-mapped credible sets that were assumed to include all the causal variants with 95% confidence (95% credible set).

To assess the functional relevance of eQTLs, we utilized 18,734 autosomal variants that had significant impacts on regulatory activity using the survey of regulatory elements (SuRE) reporter technology (van Arensbergen et al., 2019) in K562 cells. We made 10,000 null sets of variants from 5,748,566 autosomal variants that were analyzed in SuRE study while adjusting for the distance to TSS with SuRE significant variants. The enrichment score was calculated by dividing the observed overlapped number of variants by the mean of the number of overlaps in the null sets. Confidence intervals were estimated using 10,000 times bootstrapping of randomly sampled null sets with replacement.

Context-dependent eQTL analysis

We tested for the significance of the interaction between genome-wide gene expression and eQTL effects by performing a likelihood ratio test between the 2 nested models using the R anova function (two-sided). The null model, H_0, and alternative model, H_1 are detailed in the following equations:

$$H0 : E \approx I + \beta 1G + \beta 2P$$
$$H1 : E \approx I + \beta 1G + \beta 2P + \beta 3P \times G$$

where E is the normalized eGene expression, I is the intercept, G is the eVariant genotype, P is the normalized proxy gene (pGene) expression, and β_1, β_2 and β_3 are the regression coefficients.

We normalized the expression of eGenes and pGenes in a different way. The gene expression data of eGenes were normalized between samples with TMM, converted to CPM, normalized across samples using an inverse normal transform and corrected using covariates for the top 2 genetic principal components, sample collection phase, clinical diagnosis, sex and PEER factors (the same number used in the eQTL analysis). The gene expression data of pGenes were normalized between samples with TMM, converted to CPM, normalized across samples using an inverse normal transform and corrected using covariates for the top 2 genetic principal components, sample collection phase and sex, but not for the other factors because they would remove biologically meaningful diversity across donors, following the concept of the previous study (Zhernakova et al., 2017). In the analysis for estimating the relevance of disease variance (Figure S4C), pGenes were additionally corrected using clinical diagnosis as covariates.

We excluded interactions of pGenes located within 1Mbp from eVariants, as we wanted to minimize the effect of eQTL pleiotropy. We also excluded cross-mappable gene pairs (Saha and Battle, 2018) or low-mappability genes (Saha and Battle, 2018) to reduce the technical false positives. For each top eVariant, we permutated the pGene label 200 times, modeled the null distribution of *P value*s using a beta-distribution following the concept of FastQTL (Ongen et al., 2016) and calculated adjusted *P value*s of the best associated pGene. Then, in each cell subset, we adjusted for the number of top eVariants with the Benjamin-Hochberg method and calculated the threshold of adjusted P values corresponding to FDR < 0.05. For each eQTL-pGene pair, adjusted P values were converted to Z scores and we multiplied it by the sign according to the effect of pGene expression on eQTL effect sizes (+1 for magnifying interaction and −1 for dampening interaction, which we call adjusted Z score).

To facilitate interpretation, we only considered positive Z score interactions for CIRCOS visualization (Krzywinski et al., 2009) (Figure 5A). In addition, to make robust pGene clusters and eVariant clusters, we limited the analysis to pGenes that had more than 5 strong interactions (Z > 3.5), eVariants that had more than 1 strong interaction (Z > 3.5) or eVariants that had IMD GWAS variants in LD ($r^2 \geq 0.8$ in EUR and EAS). These pGenes and eVariants were each clustered into 8 groups by K-means clustering based on interaction Z scores, and strong interactions (Z > 3.5) were visualized as links (Figure 5A). For visualization of the interactions of specific GWAS loci, interactions with Z > 2.5 were linked, and in order to overview the interactions with annotated gene sets, the same set of filtered pGenes as Figure 5A were utilized (Figures 6D and 6F).

To infer events that were represented by pGenes, we compared pGenes with annotated gene sets. Interferon signature genes were defined as the union of "HALLMARK_INTERFERON_ALPHA_RESPONSE" and "HALLMARK_INTERFERON_GAMMA_RESPONSE" from MSigDB hallmark gene sets (Liberzon et al., 2015). Cell cycle associated genes were extracted from "HALLMARK_E2F_TARGETS" from MSigDB hallmark gene sets. DN2 marker genes were defined as upregulated genes in DN2 cells, as recently described (Jenks et al., 2018). Genes upregulated and downregulated in aging were defined from our dataset. After fitting the following formula with linear mixed model,

$$Expression \approx (1|Celltype) + (1|Diagnosis) + (1|Individual) + \log(Age) + \log(PSLmg + 0.1)$$

P value corresponding to the age term was calculated with the lmerTest (Kuznetsova et al., 2017) R package and genes with FDR < 0.01 were defined as up- or downregulated genes with aging. The gene sets are detailed in Table S10. The overlap *P value* of each pGene cluster and gene sets were calculated with one-sided Fisher's exact test.

Please cite this article in press as: Ota et al., Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases, Cell (2021), https://doi.org/10.1016/j.cell.2021.03.056

STAT1 and STAT2 binding motifs were evaluated using the motifbreakR (Coetzee et al., 2015) package with default parameters, and we considered SNPs with at least one allele in their LD proximities ($r^2 \geq 0.95$) achieving a motif matching p value below 1×10^{-5} with the information content method as binding motif disrupting SNPs.

For the pathway enrichment analysis of context-dependent eGenes, we compared eGenes with MSigDB immunologic signature gene sets (Godec et al., 2016). From the significantly enriched pathways, we illustrated the result with "GSE13485_CTRL_VS_DAY7_YF17D_VACCINE_PBMC_DN" (that was the most significantly enriched among all MSigDB immunologic signature gene sets), "GSE18791_UNSTIM_VS_NEWCATSLE_VIRUS_DC_6H_DN" and "GSE14000_UNSTIM_VS_16H_LPS_DC_TRANSLATED_RNA_DN" (Figure S4D). Disease variance-associated eGenes (Figure S4D) were defined as context-dependent eGenes for which the absolute value of interaction Z score dropped more than 0.5 after regressing out disease variance from the pGene expression data in any cell type. With this definition, 626 genes were classified as disease variance-associated eGenes.

LD score regression analysis

We initially fine-mapped our eQTLs with CAVIAR (Hormozdiari et al., 2014), made maximal posterior inclusion probability annotations (MaxCPP) as previously reported (Hormozdiari et al., 2018) and performed tissue-by-tissue stratified LD score regression (S-LDSC) (Bulik-Sullivan et al., 2015; Finucane et al., 2015) adjusting for functional annotation ("baseline model" provided by the developers).

To evaluate the effect of cell type-shared eQTLs, we made meta-analyzed annotation of 6 representative immune cell subset eQTLs (Mem CD4, Mem CD8, naive B, CL Mono, NK and Neu) using linear mixed model based meta-analysis (Sul et al., 2013). We utilized fixed-effect *P values* after meta-analysis and fine-mapped eQTLs with CAVIAR. Subsequently, we obtained MaxCPP of meta-analysis annotation, which we call "Meta tissue" annotation. Meta tissue annotation showed significant heritability enrichment for some traits (Figure S4E). To consider tissue-shared eQTL effects together with tissue-specific eQTL effects, we jointly regressed LD scores for each cell eQTL annotation and meta-tissue annotation together with functional annotation.

Formatted GWAS summary statistics for LDSC by developers were downloaded from https://alkesgroup.broadinstitute.org/sumstats_formatted/.

Comparison with GWAS catalog variants

Summary statistics were downloaded from the NHGRI-EBI GWAS Catalog (Buniello et al., 2019) on 04/07/2019. Variants within HLA regions were excluded from the analysis. Traits under ontology EFO0005140 were utilized as "Immune mediated diseases." Traits under ontology EFO0000589 and EFO0000319 were utilized as "Metabolic traits" and "Cardiovascular traits" respectively after manual exclusion of immune-associated traits. All variants with $p < 5.0 \times 10^{-8}$ and polymorphic in the Japanese population were pruned using PLINK (Purcell et al., 2007) based on 1000G EUR reference panel with the threshold $r^2 < 0.1$ and used for subsequent analysis.

We counted the number of eQTL variants in LD with GWAS top variants ($r^2 \geq 0.8$ in both EAS and EUR populations for ImmuNexUT eQTLs and only in EUR population for GTEx eQTLs) and assessed their relative enrichment compared to random variants. We made 100,000 sets of random variants of GWAS top variants that were adjusted for distance to TSS, minor allele frequency in the EUR population and the number of LD variants in the EUR population and polymorphic in the Japanese population. Enrichment scores were calculated by dividing the observed overlapped number of variants by the mean of the number of overlaps in the null sets.

To assess the enrichment of context-dependent eQTLs to GWAS variants, we compared them to all top eVariants in ImmuNexUT. For this purpose, we pruned context-dependent eQTLs and all top eVariants, and compared the number of variants that were in LD ($r^2 \geq 0.8$ in both EAS and EUR populations) with a pruned set of GWAS variants. *P values* were calculated using two-sided Fisher's exact test.

Colocalization analysis of eQTL and SLE GWAS

From Japanese SLE GWAS summary statistics, we analyzed 32 loci outside the HLA region that fulfilled either of the following criteria; i) It achieved genome-wide significance ($p < 5.0 \times 10^{-8}$), ii) Different population SLE GWAS achieved genome-wide significance ($p < 5.0 \times 10^{-8}$) in a nearby variant (within 500kbp) and it had the possible association ($p < 1.0 \times 10^{-4}$) in Japanese GWAS. Of these loci, lead variants of 3 loci were in strong LD ($r^2 > 0.9$) with missense variants and excluded from the colocalization analysis.

For evaluation of eQTL signal colocalization with GWAS signals, we applied coloc (Giambartolomei et al., 2014), a Bayesian framework. We tested for 500 kbp windows centered on GWAS top variants and considered *PP. H4* (posterior probability of shared causal variant) ≥ 0.8 as a significant colocalization. We also required at least 1 genome-wide significant eQTL association in the tested region. With this threshold, 18 loci showed significant colocalization with at least 1 eQTL, which were also manually confirmed.

Colocalization of loci with multiple causal variants is difficult to detect with coloc (Hormozdiari et al., 2016). Therefore, we also utilized eCAVIAR (Hormozdiari et al., 2016), a Bayesian fine-mapping approach that can deal with multiple causal variants. We tested for 100 variants centered on GWAS top variants. In their original work, the authors defined CLPP (a metrics for colocalization, that depends on the complexity of LD) ≥ 0.01 as significant colocalization threshold for eQTL and GWAS. When we compared coloc PP. H4 and eCAVIAR CLPP by rheumatoid arthritis GWAS data colocalization analysis with ImmuNexUT eQTL (Figure S4G), we observed that CLPP threshold 0.01 resulted in a relatively high frequency for false positives (we defined PP.H4 < 0.25 as true negative, and specificity = 0.81 with CLPP threshold 0.01, while specificity = 0.97 with CLPP threshold 0.03). Here our aim was to use eCAVIAR as complemental method to coloc while keeping low false positive ratio, thus we set CLPP ≥ 0.03 as a threshold for significant

Please cite this article in press as: Ota et al., Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases, Cell (2021), https://doi.org/10.1016/j.cell.2021.03.056

colocalization. eCAVIAR reproduced the same colocalization as coloc in 14 loci and revealed 2 additional colocalization loci (*CDH1* and *GPX3*).

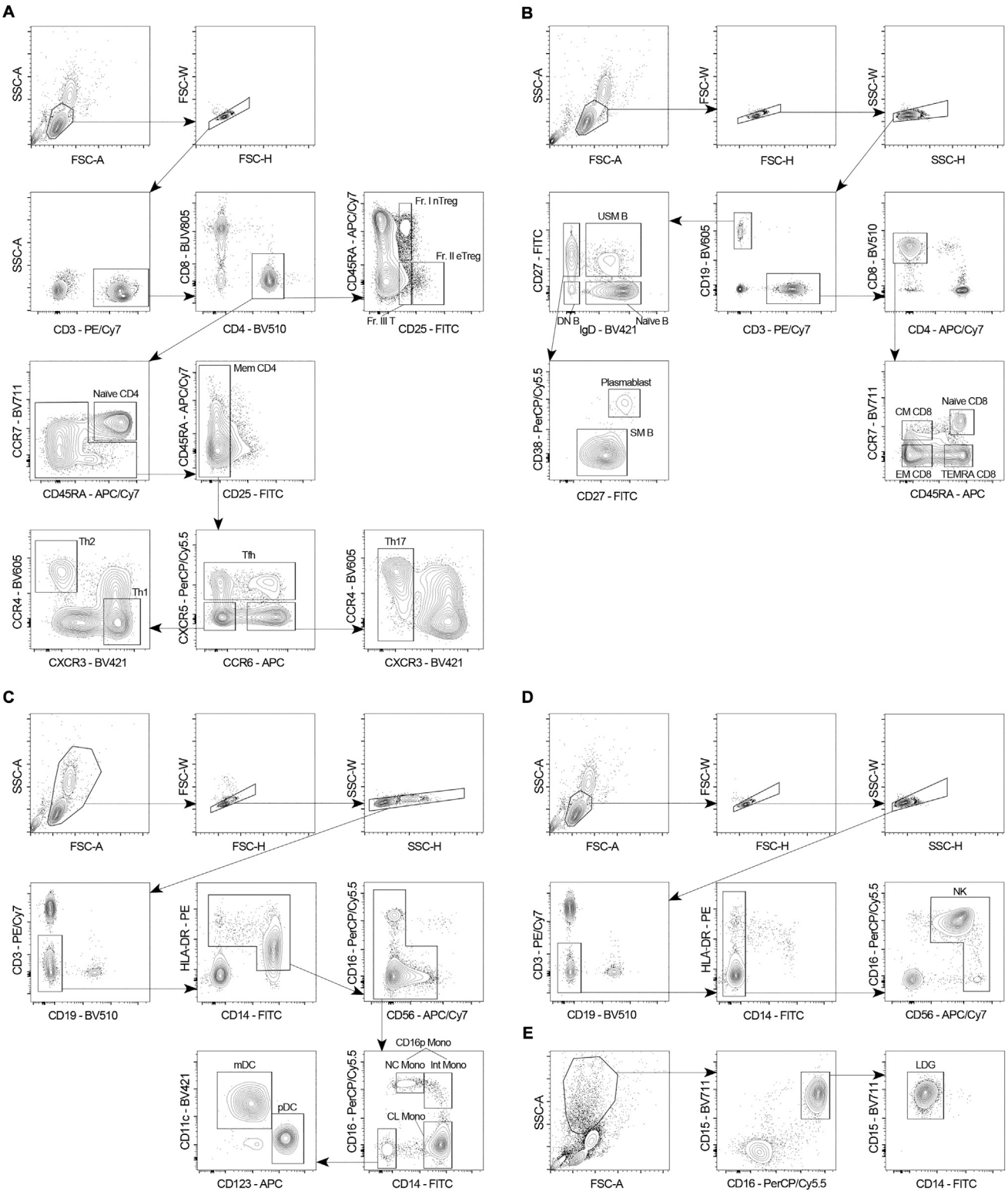
Collectively we considered 20 loci as significant colocalizations. For each locus, if 2 or more genes were colocalized, the gene with the best colocalization posterior probability (CLPP for 2 loci and *PP. H4* for the other loci) was prioritized.

For visualization, we classified colocalizations into those in the top 30th percentiles of posterior probability (corresponds to $PP.H4 \geq 0.93$ and CLPP ≥ 0.095), top 30-70th percentiles of posterior probability (corresponds to $0.93 > PP.H4 \geq 0.87$ and $0.095 >$ CLPP ≥ 0.048) and the remaining (Figure 7A). The list of colocalized genes, Japanese GWAS top variants and nearby genome-wide significant variants in different studies are summarized in Table S7.

Statistical test

The statistical test performed are indicated in the figure legends or STAR Methods. Throughout the analyses, multiple test correction was performed with Benjamini-Hochberg procedure to obtain corrected q-values unless otherwise indicated.

Supplemental figures



(legend on next page)

Figure S1. Sorting strategy for subdivided immune cell subsets, related to Figure 1

(A–E) Representative FACS plots describing gating strategies for $CD4^+$ T cells (A), B cells and $CD8^+$ T cells (B), monocytes and dendritic cells (C), NK cells (D) and LDG (E) in "phase 2" of our study (STAR Methods). See Table S8 for full description of subset names. See Table S9 for flow-cytometry antibodies used in this study.

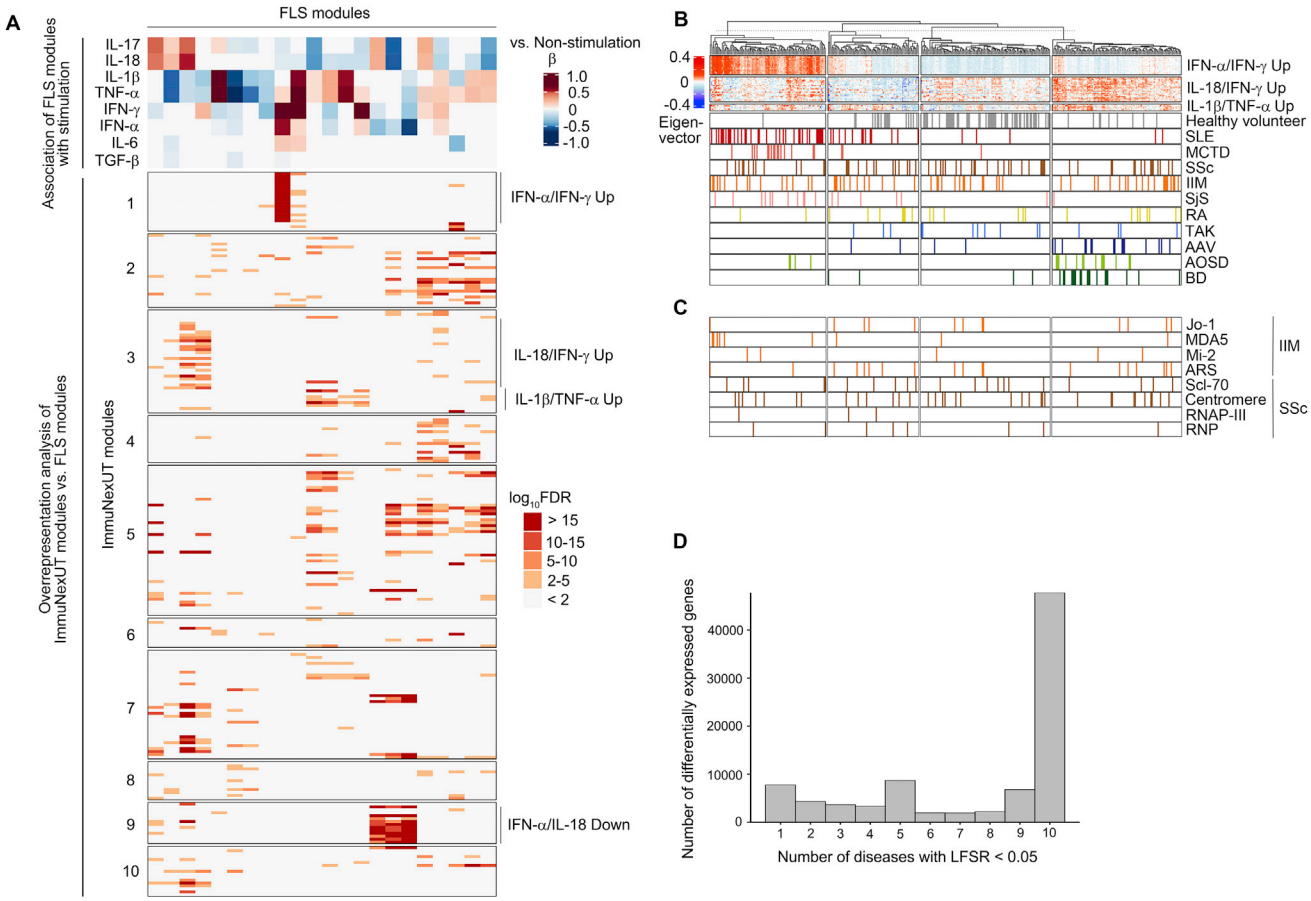


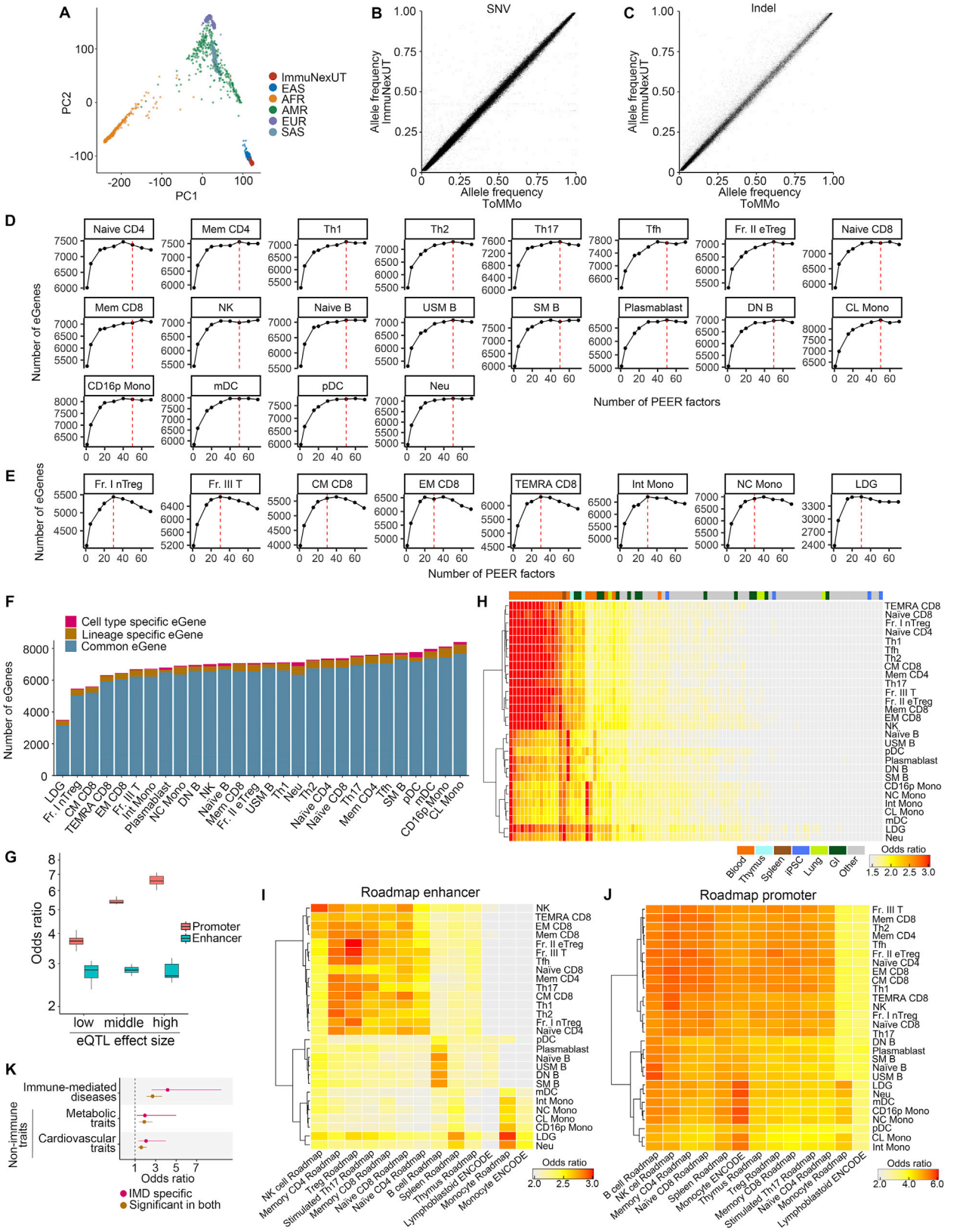
Figure S2. Characterization of gene-expression data, related to Figure 1

(A) Full results of cytokine pathway enrichment analysis of ImmuNexUT gene modules. ImmuNexUT gene modules (row in bottom panel) were compared to gene modules made from the cytokine-stimulated fibroblast-like synoviocytes (FLS) dataset (column) (STAR Methods). Upper panel shows the associations of cytokine stimulated FLS gene modules with the types of stimulation. Linear regression β value comparing the eigenvector of each module in stimulated FLS from non-stimulated FLS are plotted. In the bottom panel, ImmuNexUT gene modules (row) are arranged in the same order as Figure 1E, and the $\log_{10}$ FDR of overrepresentation analysis with FLS modules are depicted. Representative ImmuNexUT modules that showed enrichment in specific stimulation-associated FLS modules are annotated on the right.

(B) Hierarchical clustering of donors based on eigenvectors of representative gene modules. Gene modules upregulated in IMDs and annotated in Figure 1E are utilized for clustering.

(C) Distribution of autoantibody positive cases. Each row represents autoantibodies representative in IIM and SSc. MDA5, anti-melanoma differentiation-associated gene 5 antibody. ARS, anti-aminoacyl tRNA synthetase antibody. RNAP-III, anti-RNA polymerase III antibody. RNP, anti-ribonucleoprotein antibody.

(D) Distribution of the number of diseases in which genes were differentially expressed compared to healthy volunteers. The difference was considered to be significant if it had a mash local false sign rate (LFSR) of < 5%. Results of 19 cell subsets are plotted together (STAR Methods).



(legend on next page)

Figure S3. Validation of genotype data, optimization of the number of latent factors, and characterization of immune cell eQTLs, related to Figure 2

(A) Principal component analysis of ImmuNexUT genotype data together with 1000 genomes project phase 3 data. EAS, East Asian. AFR, African. AMR, Ad Mixed American. EUR, European. SAS, South Asian.

(B and C) Allele frequency comparison of SNVs (B) and indels (C) between ImmuNexUT and Tohoku Medical Megabank Organization (ToMMo) data (Tadaka et al., 2018).

(D and E) Number of significant eGenes (FDR < 0.05) in cell types with > 350 samples (D) and $\leq$ 350 samples (E) with incrementation of PEER factors (STAR Methods). Red dotted lines indicate the number of PEER factors utilized for eQTL analysis.

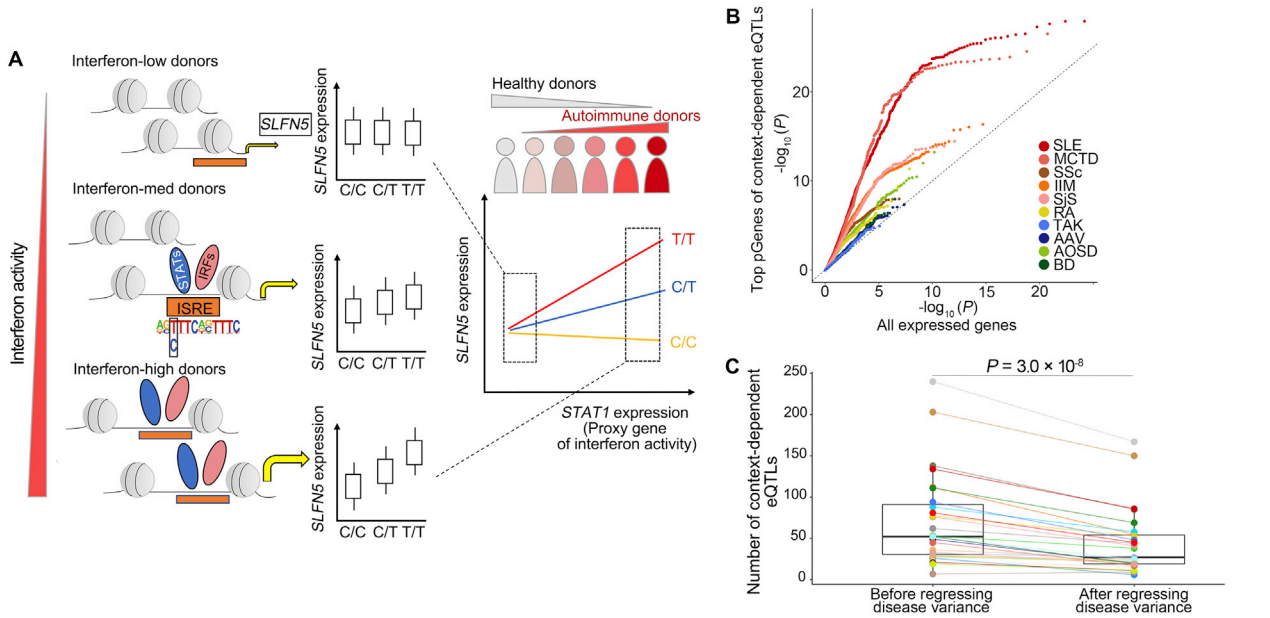
(F) The number of significant eGenes colored by their sharing by cell types or lineages. Here, each cell is classified into 1 of 3 lineages: T/NK cells, B cells and myeloid cells. When eGenes are shared among cell types in only one lineage, we classified them as "lineage specific."

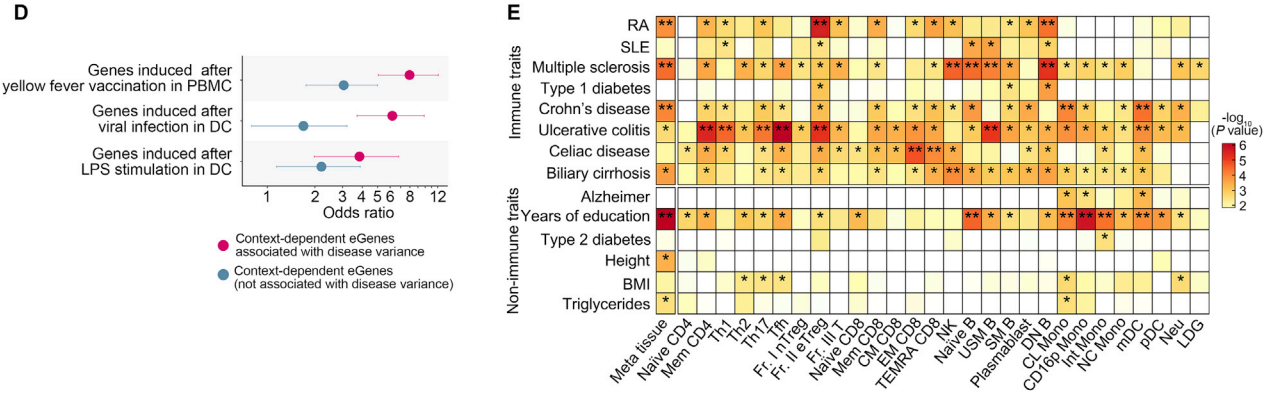
(G) Enrichment of eQTLs in promoters or enhancers of the corresponding cell types from Roadmap data (STAR Methods). Expression QTLs are stratified by absolute values of effect sizes into equal 3 groups.

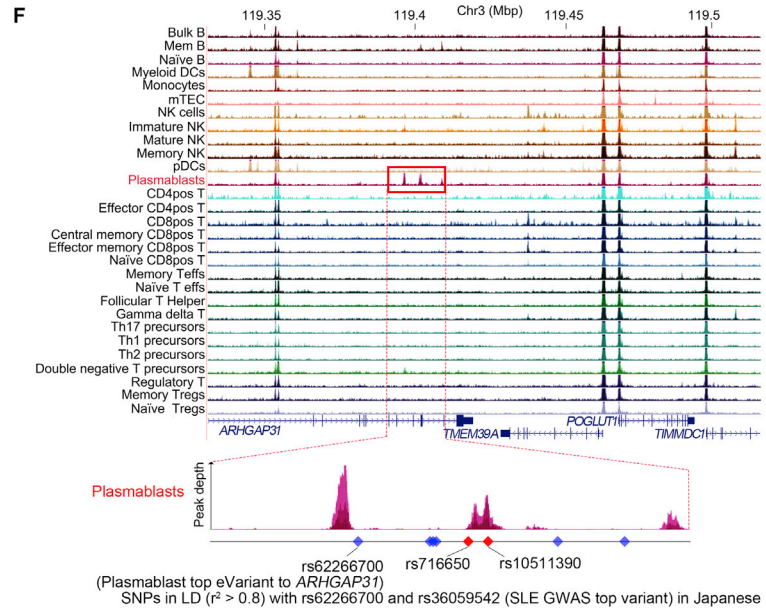
(H) Enrichment analysis of top eVariants in enhancer regions from Roadmap data. Cell types in Roadmap samples are labeled with color. Our eQTL variants showed enrichment especially in blood, thymus or spleen samples that are abundant in immune cells. GI, gastrointestinal tract.

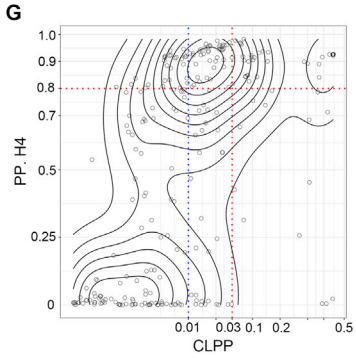
(I and J) Enrichment of top eVariants in enhancers (H) and promoters (I) of representative immune cells from Roadmap project.

(K) Enrichment of IMD-specific eQTLs and eQTLs significant in both healthy volunteers and IMD patients to GWAS top variants. Bars indicate 95% CI.









(legend on next page)

Figure S4. Characterization of context-dependent eQTL and association of ImmuNexUT with immune disease genetics, related to Figures 5, 6, and 7

(A) Scenario of context-dependent eQTLs, related to Figure 5. Interferon activates expression of some genes via epigenomic remodeling (Park et al., 2017). When chromatin is remodeled and the interferon signature response element (ISRE) is opened for transcription factor binding, the effect of the SNP in ISRE for downstream gene expression becomes evident. Here, the eQTL effect size partially depends on the fraction of remodeled cells. Consequently, the eQTL effect size is now dependent on interferon signature gene expression (e.g., *STAT1*) that reflects the interferon signature strength heterogeneity between donors. Similar eQTL effect heterogeneity can be observed when samples are heterogenous due to the cell type mixture or different degrees of various signal strengths.

(B) Comparison of differential expression *P value*s of top pGenes (the most significantly interacted proxy gene for each context-dependent eQTL), related to Figure 5. Differential expression *P value*s are calculated by comparing the gene expression between each IMD and healthy subjects. x axis reflects an empirical distribution of *P value*s from all expressed genes.

(C) The number of significant context-dependent eQTLs with or without regressing disease variance from the genome-wide gene expression data, related to Figure 5. Color of each dot indicates cell type as illustrated in Figure 1A. Two-sided Wilcoxon signed rank test.

(D) Representative result of gene set enrichment analysis of context-dependent eGenes (STAR Methods), related to Figure 5. Bars indicate 95% CI. PBMC, peripheral blood mononuclear cell. DC, dendritic cell.

(E) Association significance of meta-analysis annotation (Meta tissue, left most) or each cell type eQTL annotation with polygenic GWAS signals in tissue-by-tissue stratified LD score regression analysis, related to Figure 6A. *, FDR < 0.01; **, Bonferroni significant.

(F) Cell-type specific eQTL coincided with cell-type specific open chromatin regions, related to Figure 7B. Immune cell ATAC-seq peaks (Calderon et al., 2019) around SLE GWAS variant rs36059542 locus is shown. Only plasmablasts have open chromatin status around plasmablast specific eQTL, rs62266700. These 2 variants are in strong LD (r^2 = 0.85) in Japanese. Among 9 SNPs in LD with both of these variants, 2 SNPs (highlighted with red) coincided with ATAC-seq peaks of plasmablasts, raising the possibility of causal variants.

(G) Comparison of the metrics for colocalization outputted by coloc (*PP. H4*) and eCAVIAR (CLPP), related to Figure 7A and STAR Methods. Here we assessed the colocalization of Japanese rheumatoid arthritis GWAS top variants and ImmuNexUT eQTLs with coloc and eCAVIAR.