OUKA
The University of Osaka
Institutional Knowledge Archive

Title	Deciphering state-dependent immune features from multi-layer omics data at single-cell resolution
Author(s)	Edahiro, Ryuya; Sato, Go; Naito, Tatsuhiko et al.
Citation	Nature Genetics. 2025, 57(8), p. 1905-1921
Version Type	VoR
URL	https://hdl.handle.net/11094/102930
rights	This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

nature genetics

Article

https://doi.org/10.1038/s41588-025-02266-3

Deciphering state-dependent immune features from multi-layer omics data at single-cell resolution

Received: 18 May 2024

Accepted: 16 June 2025

Published online: 28 July 2025

Check for updates

A list of authors and their affiliations appears at the end of the paper

Current molecular quantitative trait locus catalogs are mostly at bulk resolution and centered on Europeans. Here, we constructed an immune cell atlas with single-cell transcriptomics of >1.5 million peripheral blood mononuclear cells, host genetics, plasma proteomics and gut metagenomics from 235 Japanese persons, including patients with coronavirus disease 2019 (COVID-19) and healthy individuals. We mapped germline genetic effects on gene expression within immune cell types and across cell states. We elucidated cell type- and context-specific human leukocyte antigen (HLA) and genome-wide associations with T and B cell receptor repertoires. Colocalization using dynamic genetic regulation provided better understanding of genome-wide association signals. Differential gene and protein expression analyses depicted cell type- and context-specific effects of polygenic risks. Various somatic mutations including mosaic chromosomal alterations, loss of Y chromosome and mitochondrial DNA (mtDNA) heteroplasmy were projected into single-cell resolution. We identified immune features specific to somatically mutated cells. Overall, immune cells are dynamically regulated in a cell state-dependent manner characterized with multiomic profiles.

Human omic technologies project biological mechanisms and disease pathophysiology into multi-layered matrix information with diverse resolutions. Integrative omics analysis anchored by germline genetic variants leveraged molecular quantitative trait locus (mQTL) catalogs[1]. Multi-layered mQTL catalogs synergistically answered functional annotation of the variants and filled the path from large-scale human disease genetics (that is, genome-wide association studies; GWASs) to outcome clinical phenotypes[2–10]. Such efforts initially started with bulk RNA expression profiles[2,3] and have expanded to include highly diverse layers such as proteomics[4] and metagenomics[5]. Of these, recent technological advances in single-cell RNA sequencing (scRNA-seq) have successfully elucidated cell state heterogeneity in a variety of tissues and environments[11,12]. Genetic association mapping with scRNA-seq profiling could capture continuous genetic effects along cellular states across discrete cell types and provide more granular insights into the molecular mechanisms of human trait-associated genetic variants[6–10]. Nevertheless, current single-cell expression QTL (sc-eQTL) resources are mostly centered on European ancestry[6,8,13], which rationalizes the need to construct multi-layered omics with single-cell resolution in non-European ancestries.

As a common but overlooked human mQTL layer, we additionally propose a value of using somatic genetic variations, which have been studied mainly in the field of cancer. Clonally expanded blood cells with somatic mutations or clonal hematopoiesis (CH) are common in apparently healthy individuals[14,15] and increase the risk not only for hematological malignancies[16–18] but for a variety of benign disorders, including cardiovascular disease[19,20] and infectious diseases[21,22]. However, the biological mechanisms underlying the associations between CH and benign disorders remain unknown. Refining somatic mutation spectra with finer resolution, namely, single cells, should

e-mail: r.edahiro@imed3.med.osaka-u.ac.jp; kumanogo@imed3.med.osaka-u.ac.jp; yuki-okada@m.u-tokyo.ac.jp

especially contribute to deciphering biological mechanisms and driver-or-passenger discussions[22–26], and there is definitely a growing interest in using multi-layer omics to elucidate the mechanisms of phenotypic effects of somatic events.

To interpret cell state-dependent biological phenomena by deconvoluting immune features from multi-layer human omics into single-cell resolution, here we constructed a multiomic immune cell atlas, the Osaka Atlas of Immune Cells (OASIS), from 235 Japanese persons including patients with COVID-19 and healthy individuals. The 5′ single-cell transcriptomics data profiling over 1,500,000 peripheral blood mononuclear cells (PBMCs), which are characterized by a large number of cells per sample and the inclusion of cells under in vivo immunological stimulation by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection, resulted in covering a variety of cell states. OASIS links these single-cell transcriptomics data with host genetics, plasma proteomics and metagenomics data.

We mapped the germline genetic effects on gene expression within cell types and across cell states. We also investigated HLA and genome-wide associations with variable–diversity–joining (VDJ) gene usage according to clinical status and cell types. To interpret GWAS results, we then conducted colocalization analysis of GWAS signals and our eQTLs and differential gene and protein expression analysis with polygenic risk scores (PRSs) across clinical status and cell types. In addition, by leveraging whole-genome sequencing (WGS) and SNP genotyping, we captured various somatic mutations such as mosaic chromosomal alterations (mCAs), loss of the Y chromosome (LOY) and heteroplasmy in mtDNA (mt-heteroplasmy) at single-cell resolution. We evaluated the immune features of somatically mutated cells, which were characterized as cell type-specific clonal expansions. We assessed the reactivity of clonally expanded mCAs in a patient with COVID-19 against major SARS-CoV-2 antigens using B cell receptors (BCRs). We also evaluated the effect of the gut microbiome on the differential abundance of peripheral immune cells based on a cell–cell similarity structure.

These integrative analyses at single-cell resolution demonstrated that immune cells were dynamically regulated in a cell state-dependent manner defined by the features of multi-layer omics. Our non-European, multi-layered and diversified immune single-cell atlas will contribute to equity in global diversity of human genomics and will be a valuable resource to understand complex human traits.

Results

The OASIS cohort with multi-layered single-cell omics

The OASIS cohort consists of 88 patients with COVID-19 and 147 healthy individuals of Japanese ancestry with multiomics data of immune cells (n = 235; Fig. 1a, Supplementary Fig. 1a,b and Supplementary Table 1). We performed scRNA-seq and single-cell VDJ sequencing (scVDJ-seq) on 2,059,141 PBMCs using the 10x Genomics Chromium platform and obtained 1,506,953 high-quality cells (Methods and Supplementary Fig. 1c). We manually annotated cells based on RNA expression of known marker genes[27–31]. We first defined seven major cell types (L1) according to the Azimuth L1 annotation (Fig. 1b)[27]. Next, we further identified 28 cell states (L2) and annotated ten cell types (L_{OneK1K}) for comparison with OneK1K[6] (Fig. 1b, Extended Data Fig. 1a and Supplementary Table 2). We verified our manual fine annotation using scPred[32] with the Azimuth L2 annotation as a ref. 27, which showed high concordance (Extended Data Fig. 1b).

We generated WGS data for eQTL analysis and detecting mt-heteroplasmy and SNP array data for detecting mCAs from all samples (Methods). We measured the expression of 2,925 plasma proteins with the Olink assay for 227 samples. We also performed whole-genome shotgun sequencing analysis of 131 fecal DNA samples from healthy individuals and obtained phylogenetic relative abundances.

Immune cell type-specific single-cell *cis*-eQTL mapping

To evaluate the genetic regulation of gene expression in immune cells, we first performed single-cell *cis*-eQTL analysis with WGS data using the pseudobulk approach. We tested for association between expressed genes and genetic variants located within 1 Mb of the transcription start site in each of seven major cell types (L1) and 28 fine cell types (L2). We also mapped conditionally independent *cis*-eQTLs.

In total, we identified 23,443 and 34,297 eQTLs in L1 and L2, of which 19,641 and 30,802 were primary *cis*-eQTLs (Fig. 2a and Supplementary Table 3). We detected 488–4,901 (median = 3,176) genes with significant *cis*-eQTL effects (eGenes) in L1 and 93–4,062 (median = 862) eGenes in L2, and the number of eGenes varied widely across cell types. We observed a strong association between the number of eGenes and that of cells per sample of the corresponding cell types as previously reported[6,8] (Extended Data Fig. 2a). Primary significant eQTL effect sizes were negatively correlated with the number of cells per sample of the corresponding cell types (Fig. 2b). Surprisingly, the numbers of primary eQTLs detected per cell type in this cohort were equal to or greater than those of OneK1K[6], which has about four times as many samples as our cohort (Extended Data Fig. 2b and Supplementary Table 4). Therefore, we extensively evaluated the relationship between sample size and cell counts per sample in *cis*-eQTL discovery by downsampling both at L1. The number of discovered eQTLs showed a linear decrease when downsampling the rate of cell counts per sample as well as sample sizes, and a similar relationship was observed in OneK1K (Extended Data Fig. 2c,d). These observations imply that the statistical power to detect eQTLs is highly dependent on the number of cells profiled.

Next, we evaluated how much eGenes were shared across cell types. We observed that 3,422 of 8,047 eGenes were cell type specific, whereas 1,214 eGenes were shared by more than five cell types in L1 (Fig. 2c and Supplementary Table 5). In L2, 2,613 of 7,386 eGenes were significant in only one cell type. We also compared eQTL effects among cell types and observed a high level of eQTL sharing especially within the same lineages (T and natural killer, B and myeloid cells; Fig. 2d and Supplementary Table 6). We compared the effects of significant eQTLs to the bulk eQTL dataset from Japanese individuals[33] and found high concordance (Supplementary Fig. 3 and Supplementary Table 7).

We compared *cis*-eQTLs from our cohort with those from OneK1K[6] consisting of Europeans using the multivariate adaptive shrinkage method[34] (Methods). The significant eQTLs in OneK1K were more likely to be replicated in the corresponding cell types of OASIS than those in OASIS (Extended Data Fig. 3a), which was more pronounced in cell types, with a higher number of eGenes in OASIS than in OneK1K. In naive $CD4^+$ T ($CD4_N$) cells, eQTLs that were not replicated in the other cohort showed larger differences in minor allele frequency (MAF) between East Asian and European populations than replicated eQTLs (Extended Data Fig. 3b). The direction of significant eQTL effects was almost the same in the other cohort across all cell types (Extended Data Fig. 3c). The median proportions of shared eQTLs by magnitude in the other cohort for ten cell type pairs were 34.3% for OASIS eQTLs and 69.3% for OneK1K eQTLs when the factor was set to 0.5 (Extended Data Fig. 3c and Supplementary Table 8). Similar to the replication, eQTLs in $CD4_N$ cells that were not shared between the two cohorts tended to show larger differences in MAF between East Asian and European populations than shared eQTLs (Extended Data Fig. 3d). These results suggest the importance of constructing sc-eQTL resources from different populations with larger sample sizes.

We annotated significant eQTLs using chromatin state predictions from eight immune cell types of the Roadmap Epigenomics project[35]. Primary eQTLs were enriched in both promoter and enhancer regions of Roadmap immune cells. However, enrichment of eQTLs in promoter regions was shared among cell types, whereas that in enhancer regions was more cell type specific (Fig. 2e and Supplementary Table 9). When primary eQTLs were stratified by effect sizes,

https://doi.org/10.1038/s41588-025-02266-3

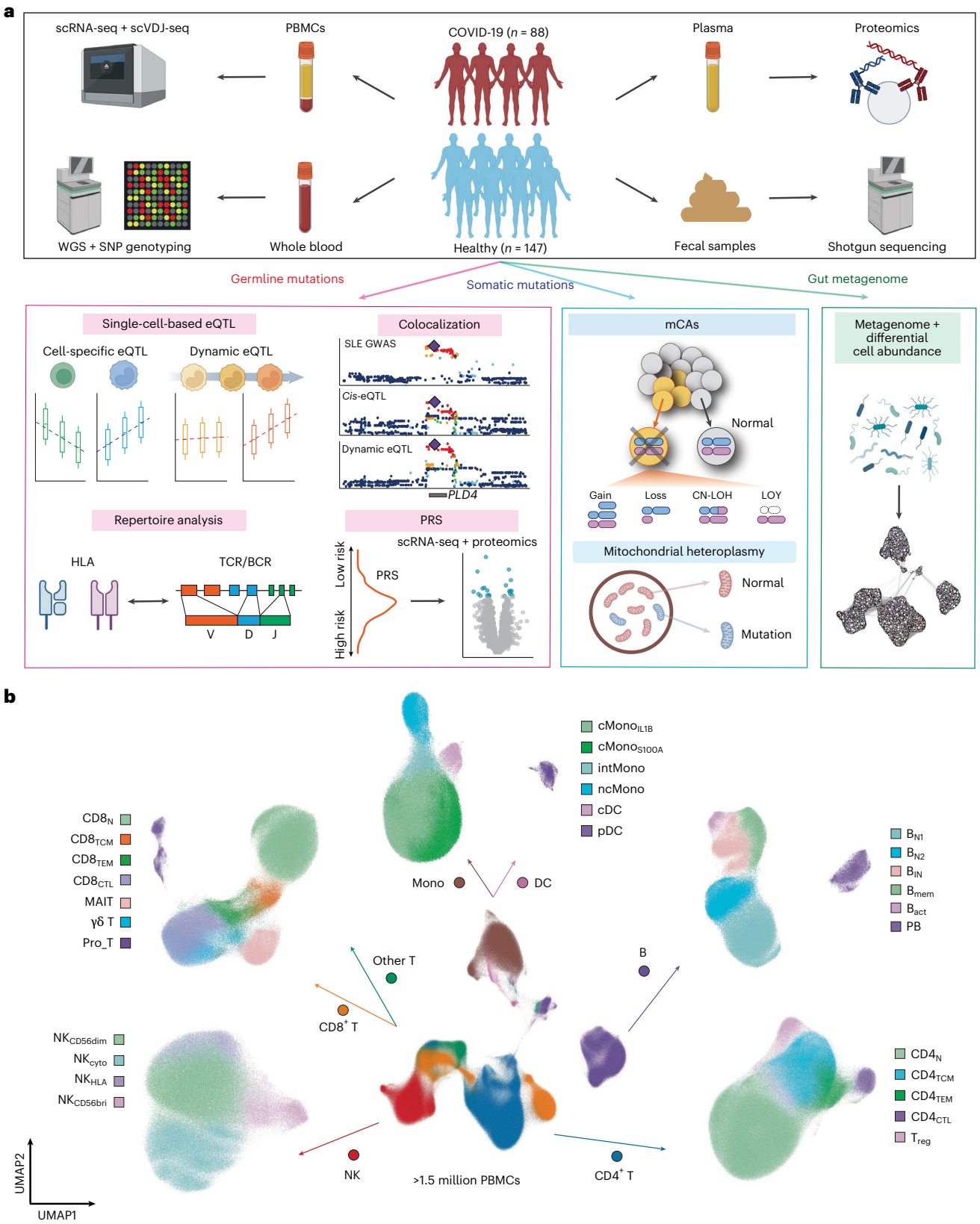


Fig. 1 | Overview and scRNA-seq data of OASIS. a, Overview of the study design. SLE, systemic lupus erythematosus. **b**, UMAP embedding of scRNA-seq data for 1,506,953 cells. Seven major cell types and 28 fine cell types were defined by RNA expression of marker genes (Extended Data Fig. 1a). B_{act}, activated B cell; B_{mem}, memory B cell; B_{N1}, type 1 naive B cell; B_{N2}, type 2 naive B cell; BIN, intermediate B cell; $CD4_{CTL}$, $CD4^+$ cytotoxic T celle; $CD4_{TCM}$, $CD4^+$ central memory T cell; $CD4_{TEM}$, $CD4^+$ effector memory T cell; $CD8_{CTL}$, $CD8^+$ cytotoxic T cell; $CD8_N$, naive $CD8^+$ T cell; $CD8_{TCM}$, $CD8^+$ central memory T cell; $CD8_{TEM}$, $CD8^+$ effector memory T cell; cDC, conventional dendritic cell; cMono, classical monocyte; cMonoIL1B, IL1B cMono; $cMono_{S100A}$, S100A cMono; intMono, intermediate monocyte; MAIT, mucosal-associated invariant T cell; Mono, monocyte; ncMono, non-classical monocyte; NK, natural killer cell; NK_{cyto}, cytokine NK cell; pDC, plasmacytoid dendritic cell; Pro_T, proliferating T cell; T_{reg}, regulatory T cell. Panel a created with BioRender.com.

https://doi.org/10.1038/s41588-025-02266-3

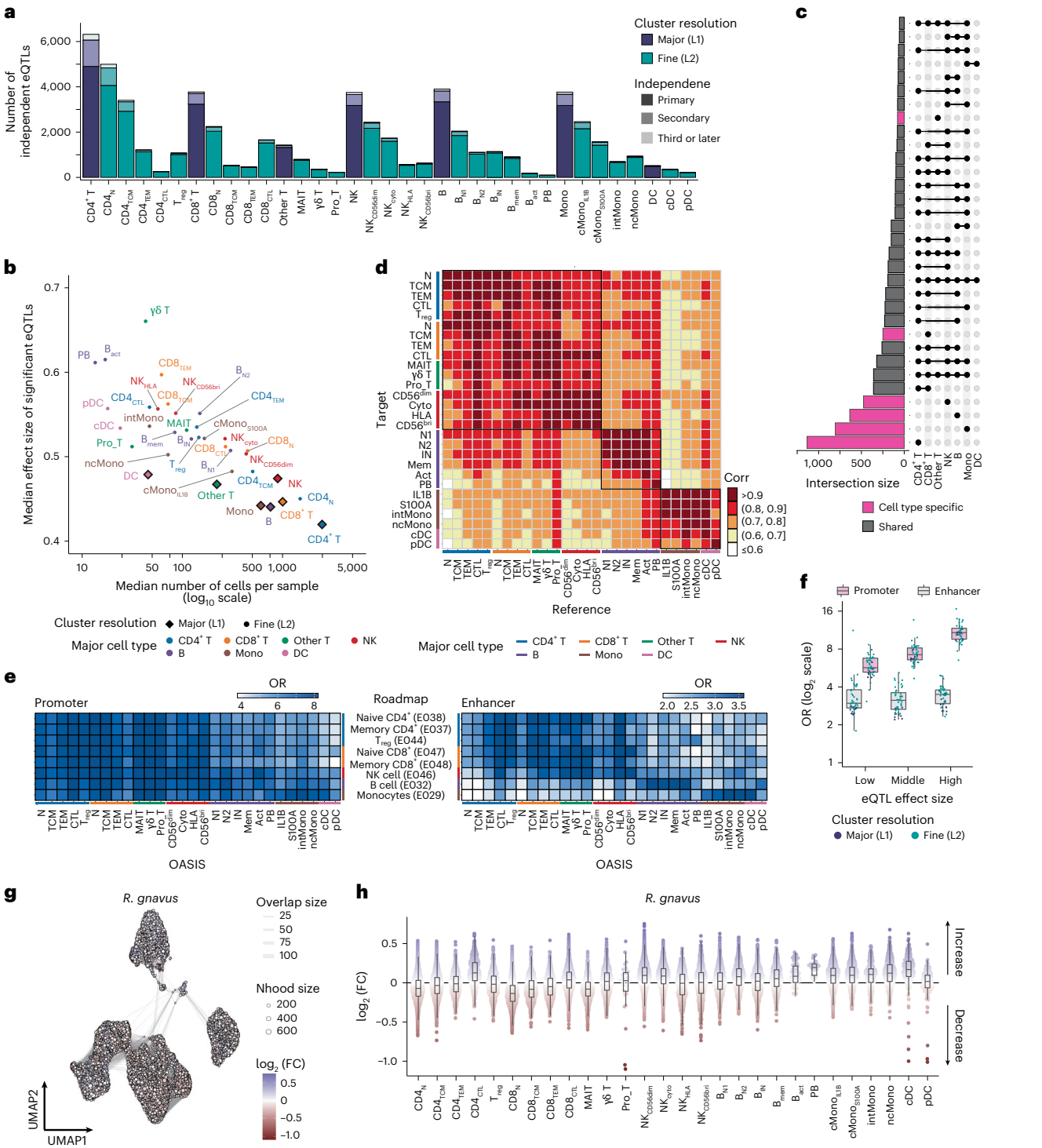


Fig. 2 | Germline genetic variants and the gut metagenome affect transcriptional and phenotypic profiles of immune cells. a, Number of significant independent eQTLs in each cell type of L1 and L2 levels. The bars are colored by cluster resolution (L1 or L2 level) and shaded by independence of eQTL mapping. DC, dendritic cell. **b**, Scatterplot depicting the correlation between median number of cells per sample and median effect size of significant eQTLs in each cell type. The *x* axis is on a $\log_{10}$ scale. **c**, Number of significant eGenes shared among cell types in the L1 level. eGenes identified in only one cell type are colored pink, and those in multiple cell types are connected by lines and colored gray. **d**, Heatmap showing a pairwise comparison of eQTL effect size. Only significant eQTLs in one cell type (reference) that could be evaluated in other cell types (target) were analyzed. **e**, Heatmap depicting the enrichment of primary significant eQTLs of L2 cell types in promoter or enhancer regions of eight representative immune cells from the Roadmap project. **f**, Box plot showing the enrichment of primary eQTLs in promoters or enhancers of the corresponding cell types ($n = 28$) from the Roadmap project, according to effect sizes (Methods). **g**, Graphical representation of neighborhoods identified by Milo in healthy individuals ($n = 131$). Nodes are neighborhoods, colored by $\log_2$ (fold change (FC)) by the relative abundance of *R. gnavus* adjusted for age, sex and sequencing groups. Sizes correspond to the number of cells in a neighborhood. Graph edges depict the number of cells shared between adjacent neighborhoods. Nhood, neighborhood. **h**, Beeswarm plot and box plot showing the distribution of adjusted $\log_2$ (FC) in neighborhoods ($n = 43,089$) by the relative abundance of *R. gnavus* among L2 cell types. Colors are represented in the same way as in **g**. Boxes denote the interquartile range (IQR), medians are shown as horizontal bars, and whiskers extend to 1.5 times the IQR in **f**,**h**. N, naive; TCM, central memory T cell; TEM, effector memory T cell; CTL, cytotoxic T cell; N1, type 1 naive; N2, type 2 naive; IN, intermediate; Cyto, cytokine; Act, activated; corr, correlation.

promoter enrichment was more predominant at larger effect sizes of eQTLs, but enhancer enrichment was similar regardless of effect sizes (Fig. 2f). Considering that more than 60% of causal variants in autoimmune diseases map to immune cell enhancers[36], this result motivates us to identify more eQTLs in enhancer regions that do not necessarily have large effect sizes.

Gut bacterial abundance is associated with immune cell abundance

To reveal the sets of immune cells that were differentially abundant with bacterial abundance at finer resolution, we performed differential abundance analysis using single-cell neighborhoods (Milo)[37]. In this analysis, we focused on the following three species, all of which had significant associations with more than two cell types in the proportional analysis (Supplementary Fig. 5a) and have been reported to be involved in human diseases: *Ruminococcus gnavus*[38], *Prevotella copri*[39] and *Bacteroides vulgatus*[40]. We identified 43,089 neighborhoods among PBMCs, none of which showed significant differential abundance with *R. gnavus* abundance (Fig. 2g). However, the increase in *R. gnavus* was accompanied by an increase of rare cell types (that is, $CD4^+$ cytotoxic T cells, activated B cells and plasmablasts (PBs)), along with an increase in cell types of the myeloid cluster (Fig. 2h). Similarly, differential abundance analysis revealed that the increase in *P. copri* was associated with an increase in PBs and plasmacytoid dendritic cells and a decrease in $CD4^+$ cytotoxic T cells (Supplementary Fig. 5b,c). Such differential abundances were consistent with results of the proportional analysis, demonstrating robustness. These findings highlight that integrative omics analyses anchored by single-cell data allow us to detect biological phenomena that could not be captured by conventional bulk analyses.

HLA and genome-wide associations with immune repertoires

We performed scVDJ-seq to construct a catalog of T cell receptor (TCR) and BCR repertoires, including usage of each gene across different cell types (Supplementary Fig. 6). We explored the relationship between HLA and the TCR repertoire, characterized by specific cell types. We assessed associations between HLA amino acid variants and TCR V gene usage, including TRAV and TRBV. We identified significant associations between TRAV genes and HLA amino acid variants both in HLA class I and II genes. Specifically, TRAV genes were associated with HLA class I and II gene variants exclusively in $CD8^+$ T and $CD4_N$ cells, respectively ($P < 1.1 \times 10^{-5}$; Fig. 3a,b). These results align with the well-known pattern of HLA restriction on TCRs according to T cell subpopulations[41,42]. Furthermore, in comparisons among $CD4^+$ T cell subpopulations, the associations between TRAV genes and HLA class II gene variants appeared stronger in $CD4_N$ cells than in entire (that is, $CD4^+$ T) and central memory $CD4^+$ T cells (Fig. 3b). This might reflect the stronger HLA restriction on the TCR repertoire formed at the central level (that is, in the thymus) than that formed at the peripheral level[43].

We explored the impact of HLA variation on the TCR repertoire in response to COVID-19. We observed significant interactions, particularly those for HLA class I gene variants and TRAV gene usage in $CD8^+$ T cells ($P < 2.1 \times 10^{-6}$; Fig. 3c). This may reflect the generation of different TCR repertoire in SARS-CoV-2 infection according to HLA class I variation in $CD8^+$ T cells.

Additionally, we investigated the association of genome-wide variants with the TCR repertoire, focusing on integrative features representing repertoire variation (Methods). Furthermore, we extended this analysis to TCR D and J genes, as the genetic association with these genes has been underexplored. We identified a significant association between class II HLA loci with principal component (PC)2 for TRAV gene usage in $CD4^+$ T cells as well as PC1 for TRAV gene usage in $CD4_N$ cells ($P < 2.0 \times 10^{-9}$; Fig. 3d, Extended Data Fig. 4 and Supplementary Table 11).

Lastly, we conducted the same repertoire feature-wide association analysis for BCRs. We detected significant associations between PCs for several immunoglobulin gene usages and variants located within these genes ($P < 1.3 \times 10^{-9}$; Fig. 3e, Extended Data Fig. 4 and Supplementary Table 12). Furthermore, somatic hypermutation (SHM)-related features for K and L chains were significantly associated with variants in the IGKV and IGLV genes (Fig. 3e and Extended Data Fig. 4). The stronger associations in naive B cells may be explained by the observation that SHMs in memory cells are more likely to be shaped by exposures, which can weaken the genetic effects.

Dynamic eQTL effects across two immune-related gene modules

To evaluate the dynamic genetic regulation of gene expression along continuous cell states, we investigated the dynamic effects of eQTLs in the myeloid cluster across the two gene modules related to immunity and COVID-19 severity[12,30,44] (Fig. 4a). Module 1 was highly correlated with uniform manifold approximation and projection (UMAP)1 and module 2 with UMAP2, representing different cell states. To model dynamic eQTLs, the gene modules were divided into ten bins, and the average expression profiles per individual in each bin were reconstructed for each module (Fig. 4a and the Methods). We evaluated dynamic eQTLs by testing the interaction between genotype and quantile rank using both linear and quadratic models for the two modules, respectively[6,8]. We identified 530 and 568 genes with dynamic eQTL effects (deGenes) across modules 1 and 2 from robust candidates, respectively (Methods). Of these deGenes, 352 (66.4%) and 393 (69.1%) showed dynamic eQTL effects in both linear and quadratic models, and 134 (25.3%) and 117 (20.6%) did only in the quadratic model for modules 1 and 2 (Extended Data Fig. 5a). Moreover, more than half of deGenes (324 in module 1 and 362 in module 2) were module specific (Fig. 4b), and the enriched pathways of module-specific deGenes were related to innate immunity in module 1 and antigen presentation in module 2 (Fig. 4c). We confirmed that the model was well calibrated by 1,000 permutations (Extended Data Fig. 5b and the Methods).

We annotated dynamic eQTLs using the Roadmap data[35] as we did for *cis*-eQTLs. Dynamic eQTLs showed monocyte-specific enrichment in both promoters and enhancers (Fig. 4d and Supplementary Table 16) but distinctly from that of *cis*-eQTLs (Fig. 2e). Next, we compared enrichment of dynamic eQTLs and *cis*-eQTLs in monocytes (L1) in promoters and enhancers of monocytes. In both modules, dynamic eQTLs were more enriched for enhancers and less for promoters than *cis*-eQTLs (Fig. 4e and Supplementary Table 17). Similar results were observed in the comparison of dynamic eQTLs and *cis*-eQTLs in subclusters (L2) of the myeloid cluster (for example, plasmacytoid dendritic cells; Extended Data Fig. 5c). We investigated deGenes for which dynamic eQTLs were located in functional regions, but *cis*-eQTLs of any cell type of the myeloid cluster were not (Methods). Pathway enrichment analysis of such deGenes across modules and functional categories indicated that pathways related to Toll-like receptor were enriched in module 1 and those related to antigen presentation were enriched in module 2, and their enrichment pattern varied between promoters and enhancers within each module (Fig. 4f).

Modeling dynamic eQTLs at single-cell resolution

We performed dynamic eQTL analysis at single-cell resolution (Fig. 4a and the Methods). As an example, we observed the strong cell state-dependent eQTL effect of rs11080327 for *SLFN5* specifically at specific states within classical monocytes (cMono; Fig. 4g), capturing the gene regulatory mechanism with higher resolution than previously reported[7]. Among 15 hPCs, hPC14, which represented a cell state related to the type 1 interferon pathway, showed the most significant cell state interaction (Fig. 4g and Extended Data Fig. 5d), consistent with results in the dynamic eQTL analysis with the pseudobulk approach. Similarly, eQTLs for *NFKBIZ* and *IFITM2* (rs9818678, rs741738) exhibited

https://doi.org/10.1038/s41588-025-02266-3

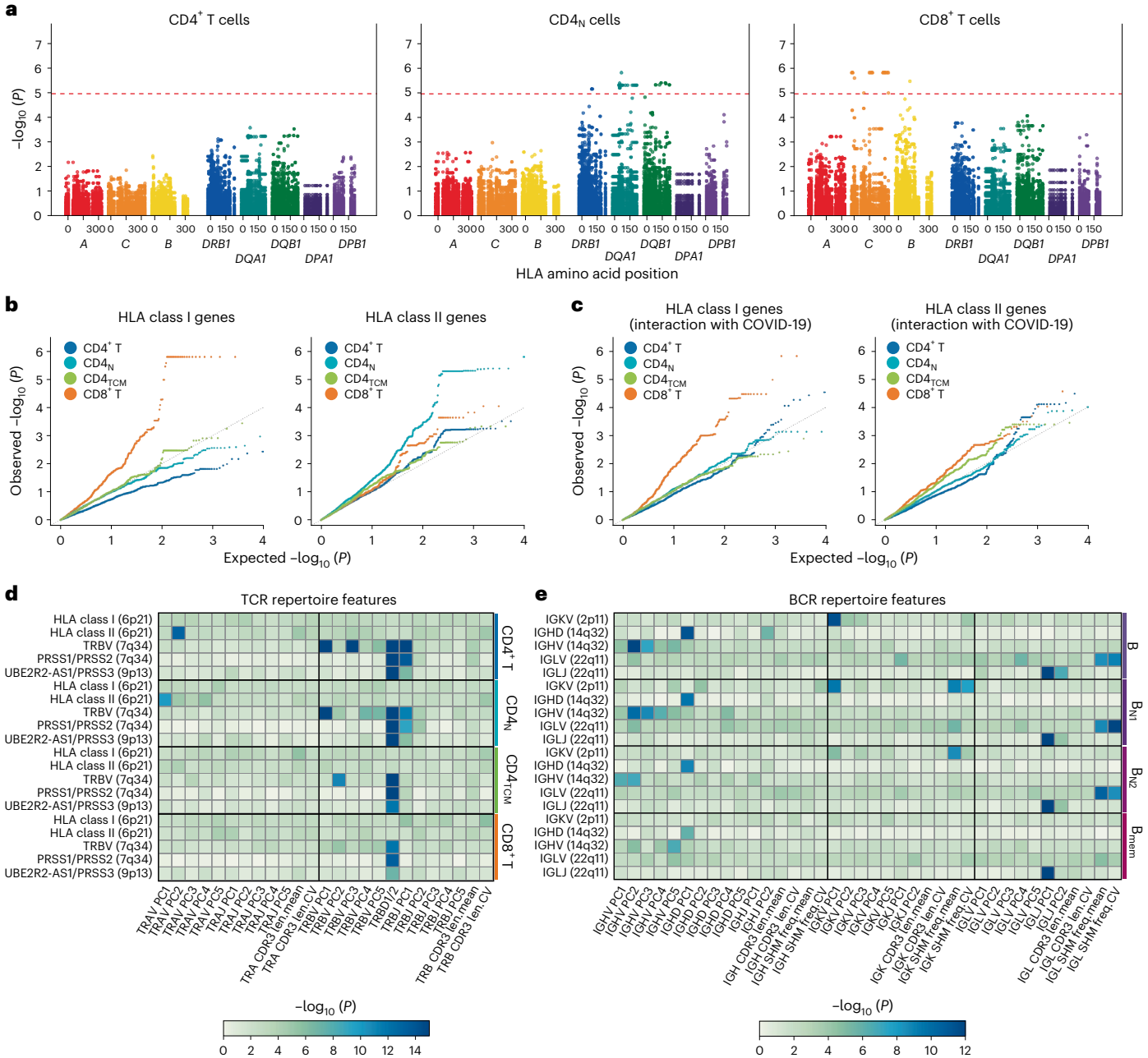


Fig. 3 | HLA and genome-wide association analysis with TCR and BCR repertoires. a, Regional plots for HLA association with TRAV gene usage in CD4$^+$ T (left), CD4$_N$ (center) and CD8$^+$ T (right) cells. In each plot, $-\log_{10}$ (*P* values) for the association tests between amino acid variants of each HLA gene and all tested TRAV gene usages are shown along with the horizontal axis representing amino acid positions. The dashed red horizontal line represents the study-wide *P*-value significance threshold. **b**,**c**, Quantile–quantile plots for the association (**b**) and interaction (**c**) tests between HLA amino acid variants and TRAV gene usage in HLA class I (left) and II (right) genes in different cell types. Vertical and horizontal axes indicate the observed and expected $-\log_{10}$ (*P* values) for the tests. **d**,**e**, Heatmaps show the maximum values of $-\log_{10}$ (*P* values) for the association tests for repertoire features (horizontal axes) within individual loci (vertical axes) for TCR (**d**) and BCR (**e**) repertoires. The position range for each locus is determined by the gene body for individual genes and the uppermost and lowermost gene bodies for multiple genes or gene clusters. Only loci that exhibited significant associations with a feature in at least one cell type are displayed. Uncorrected *P* values from two-sided tests are shown in **a**–**e**. CV, coefficient of variation; freq, frequency; len, length.

significant cell state dependence, but their cell state interactions were different (Fig. 4g). Expanding eQTL mapping into single-cell resolution should yield a more granular picture of dynamic genetic regulation.

Colocalization of GWAS variants and sc-eQTLs

To better understand the genetic regulatory mechanisms of GWAS loci, we assessed colocalization of GWAS signals from 13 complex traits in an East Asian population and our eQTL signals mapped per cell types in L1 and L2. We discovered GWAS–eQTL colocalization events (PP.H4 (posterior probability of shared causal variant) > 0.8) at 121 GWAS loci and prioritized 179 candidate trait-associated genes (Fig. 5a and Supplementary Table 19). About half of these GWAS loci (55 of 121) showed colocalization within only one major cell type (Fig. 5a), and most prioritized trait-associated genes were trait specific (Extended Data Fig. 6a). Cell types that exhibited colocalization were prominently specific to the traits (Fig. 5a and Extended Data Fig. 6b).

We next jointly colocalized dynamic eQTLs with GWAS loci. As an example, rs2841281, the lead SNP for systemic lupus erythematosus

GWAS in the *PLD4* locus, had dynamic eQTL effect only with module 1 (Fig. 5b). While this variant moderately colocalized with some *cis*-eQTLs per cell type, it colocalized more strongly with dynamic eQTLs in two bins of module 1 (posterior probability of shared causal variant (PP.H4) = 0.95 for bin 6, 0.92 for bin 7; Fig. 5b,c). In addition, single-cell-resolution eQTL modeling revealed the strongest cell state-dependent eQTL effects in the boundary region between IL1B classical monocytes ($cMono_{IL1B}$) and S100A classical monocytes (Fig. 5d). As another example, rs2836884, the lead SNP for ulcerative colitis GWAS in the *ETS2* locus, had a more dynamic eQTL effect with module 1 than module 2 (Fig. 5e). This variant showed nearly perfect colocalization with dynamic eQTLs in bin 6 of module 1 ($PP.H_4$ = 0.99; Fig. 5e,f) and had heterogeneous cell state-dependent eQTL effects even within the cell types exhibiting significant colocalization (Fig. 5g). These observations suggest the importance of considering dynamic eQTLs to interpret GWAS signals.

Context- and cell type-specific effects of polygenic risks

Beyond single variants, we exhibited dynamic gene regulation in a genome-wide manner. We further studied how PRSs affect transcriptomic and proteomic profiles across clinical status and/or cell types. We constructed PRSs of hospitalized patients with COVID-19 using PRS-CSx[45], combining the summary statistics of hospitalized COVID-19 in Europeans[46] and those in Japanese individuals[47,48] as the training data (Methods). The phenotypic variance (Nagelkerke's R^2) explained by our PRS in OASIS was 4.1%, confirming that the PRS had reasonable accuracy. We divided this PRS into four quantiles and performed differential gene and protein expression analysis with quantiles of PRSs across clinical status and cell types. We identified differentially expressed genes (DEGs) in monocytes (n = 21) and $CD8^+$ T cells (n = 2) from COVID-19 scRNA-seq and differentially expressed proteins (DEPs) (n = 184) from the COVID-19 proteome (P < 0.045 by 1,000 permutations), whereas no DEGs or DEPs were identified in healthy individuals (Fig. 5h and Extended Data Fig. 7a,b). We confirmed that DEGs and DEPs specific to patients did not directly reflect the DEGs and DEPs between patients and controls (Extended Data Fig. 7c,d). We also investigated the effect of PRSs derived from other trait GWASs and found that their effects were context and cell type specific (Extended Data Fig. 7e). These findings indicate that PRSs, like single germline variants, might affect transcriptional and proteomic profiles in a context-specific and/or cell type-specific manner.

Single-cell profiling of mCAs

To expand our mQTL catalogs to somatic genetic variations, we conducted single-cell deconvolution of a variety of somatic events detected with genomics data (Fig. 6a). Using SNP array data[14,49], we detected eight copy number alterations (CNAs) and six copy-neutral losses of heterozygosity (CN-LOHs) in seven and six samples, respectively (Supplementary Table 20). By using the individuals' mCAs as prior information and applying Numbat[50] to the scRNA-seq data, we could distinguish the mutant cells from their wild-type counterparts in each sample, except for one patient with COVID-19 containing a shorter CNA (CH12; Supplementary Table 20). Because the detection of clone cells with CN-LOH depends on germline SNP alleles embedded in the scRNA-seq raw reads, an increase of such scRNA-seq-based SNP information improves the sensitivity. We thus performed deep and long sequencing by scaling target depths from 20,000 to 100,000 reads per cell and extending a read 2 length from 90 to 270 bp. Consequently, we profiled 1.8-fold SNPs and identified 2.1-fold more mutated cells with CN-LOH than with normal sequencing conditions (Extended Data Fig. 8a). While cells with CN-LOH were moderately enriched in other T cells, strong cell type specificity of cells with CNA was observed in two patients with COVID-19 (CH01 and CH05; Fig. 6b, Extended Data Fig. 8b and Supplementary Table 21). These two patients also showed relatively larger fractions of mutant cells (median of 12.3% for three CNAs in CH01 and CH05 versus 3.06% for the other four CNAs; Extended Data Fig. 8a).

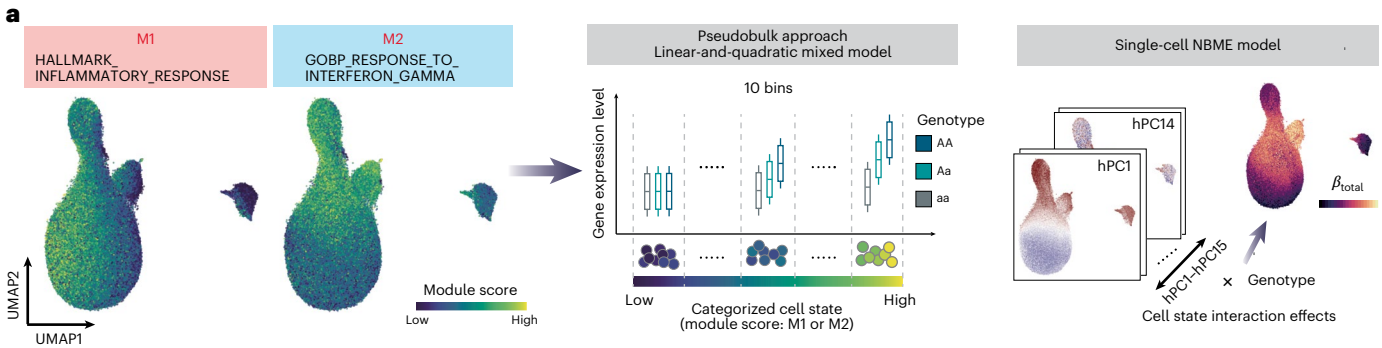
In CH01, two clones with different CNAs were enriched in monocytes (1p loss, odds ratio (OR) = 6.6 for monocytes; 15q gain, OR = 7.2 for monocytes; Fig. 6b,c). To characterize these clones, we assessed DEGs between mutant and normal cells in monocytes (Extended Data Fig. 8c). Most of the downregulated genes in monocytes with 1p loss and the upregulated genes in monocytes with 15q gain were located within the altered chromosomal regions themselves (that is, *cis*), highlighting accurate detection of mutant cells. But some showed *trans*-chromosomal DEGs including upregulation of *TNFAIP3*, one of the elements of the COVID-19-specific immune response[51], in monocytes with 1p loss (Extended Data Fig. 8c). We further evaluated the enriched pathways of DEGs between mutant and normal clones and found that immune-related pathways, such as positive regulation of cytokine production, were significant in mutant clones with 1p loss (Fig. 6d and Extended Data Fig. 8d,e).

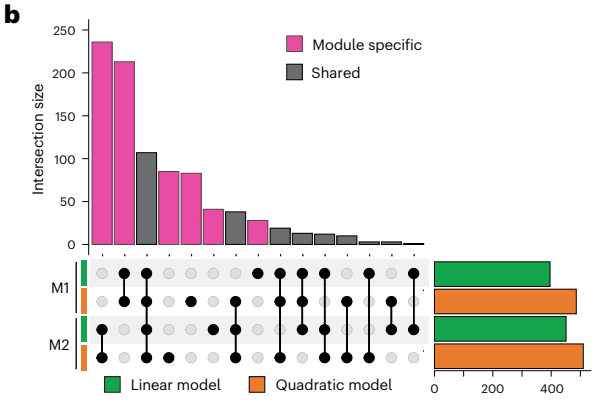
Mutant cells with 17q gain were strongly enriched in B cells in CH05 (OR = 350; Fig. 6b,e). Pathway enrichment analysis of DEGs between mutant and normal B cells showed upregulation of immune-related pathways and downregulation of response to steroid hormones in mutant cells (Extended Data Fig. 8c,f,g). This suggests that these mutant cells may reduce the effectiveness of systemic corticosteroid therapy and result in a worse prognosis. Furthermore, our integrated analysis of scVDJ-seq data identified a considerably large BCR clonotype with 17q gain in CH05 (clone size = 650; Fig. 6f). This BCR clonotype was the largest clonotype among all the samples (median of maximum clone size in each sample = 5) and was mainly composed of naive B cells, which differed from most expanded clonotypes composed of class-switched PBs (Extended Data Fig. 9a). Using recombinant antibodies derived from this expanded BCR clonotype,

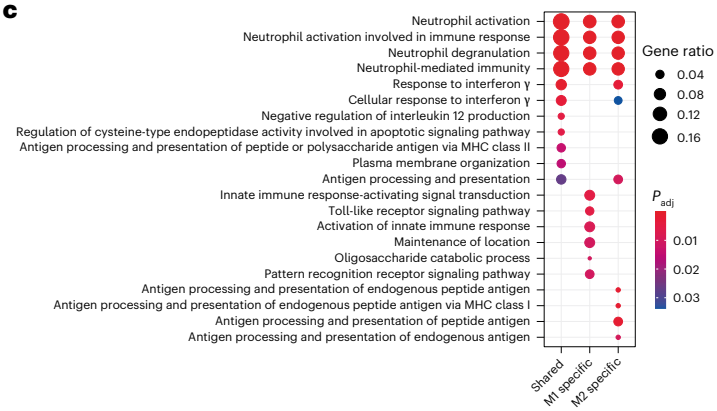
Fig. 4 | Dynamic eQTL analysis in the myeloid cluster. a, Overview of dynamic eQTL analysis. We calculated the two module scores using a gene set termed 'HALLMARK_INFLAMMATORY_RESPONSE' and 'GOBP_RESPONSE_TO_INTERFERON_GAMMA' (GO:0034341), respectively (left). Cells were split into ten windows of equal cell numbers according to each module score (middle). The figure design is based on a previous report[8]. A linear-and-quadratic mixed model was applied to test for an interaction between genotypes and module scores by the pseudobulk approach. The single-cell negative binomial mixed-effect (NBME) model was used to identify cell state-dependent regulatory effects (right). Module 1, M1; module 2, M2. **b**, Number of eGenes with a significant genotype–module interaction (that is, dynamic eGenes) in a linear or quadratic mixed model for the two modules. **c**, Top ten biological processes by gene ontology (GO) enriched in dynamic eGenes are shown for each test. MHC, major histocompatibility antigen; P_{adj}, adjusted P value. **d**, Heatmap depicting the enrichment of dynamic eQTLs in promoter or enhancer regions of eight representative immune cells from the Roadmap project for each combination of module and analysis model. **e**, Forest plots showing ORs of the overlap of dynamic eQTLs (n = 396–510; Supplementary Table 15) with functional regions in monocytes from Roadmap compared to *cis*-eQTLs in monocytes at L1 (n = 3,175) for each combination of functional region and analysis model in two modules. Dots represent ORs, and bars represent 95% confidence intervals (CIs). **f**, Top four biological processes by GO enriched in dynamic eGenes, where dynamic eQTLs, but not cis-eQTLs, are located in functional regions of monocytes from Roadmap, are shown for each module-function combination (four tests in total). **g**, UMAPs represent the cell state-dependent eQTL strength (β_{total}) for each cell calculated as a sum of the effect sizes of genotype and genotype (G) × harmonized PCs (hPCs). Labeled P values are derived from NBME and pseudobulk analysis. Heatmap showing P values for each genotype–hPC interaction in the full model of NBME analysis. Two-sided P values are uncorrected in **e**,**g**. Dot color indicates statistical significance of the enrichment (adjusted P values via the Benjamini–Hochberg method), and dot size represents the gene ratio assigned to each term in **c**,**f**.

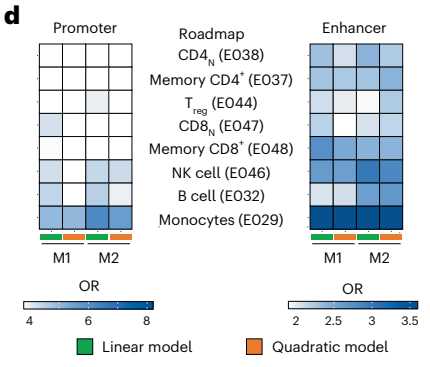
https://doi.org/10.1038/s41588-025-02266-3

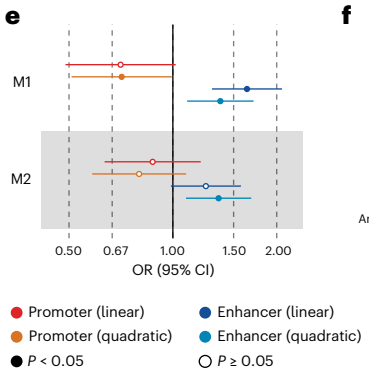
we assessed their reactivity against SARS-CoV-2 major antigens. Of interest, the BCR clonotype did not react with any tested antigens (Fig. 6g and Extended Data Fig. 9b), not supporting clonal expansion through normal antibody response to SARS-CoV-2 infection. These clonally expanded mutant B cells may have reduced diversity of the BCR repertoire in SARS-CoV-2 infection, which could negatively impact the antibody response of this patient. Our deconvolution of mCAs at single-cell resolution has the potential to elucidate the impacts of somatic mutations on immune cell functions and the development of severe infections.

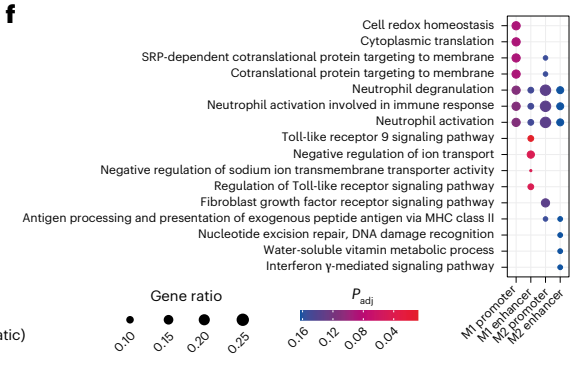


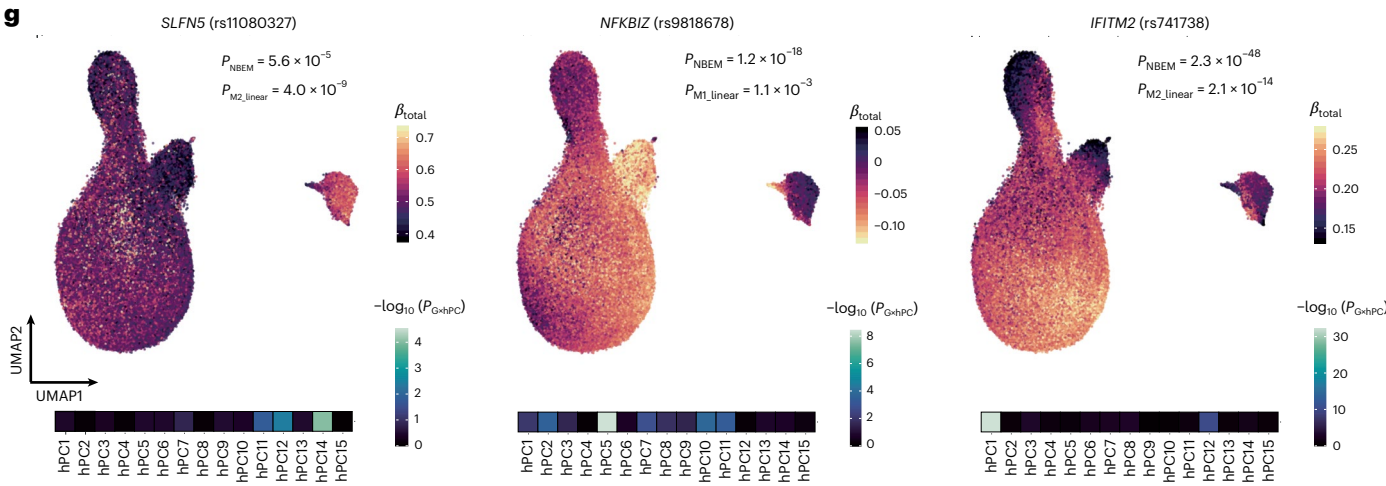












https://doi.org/10.1038/s41588-025-02266-3

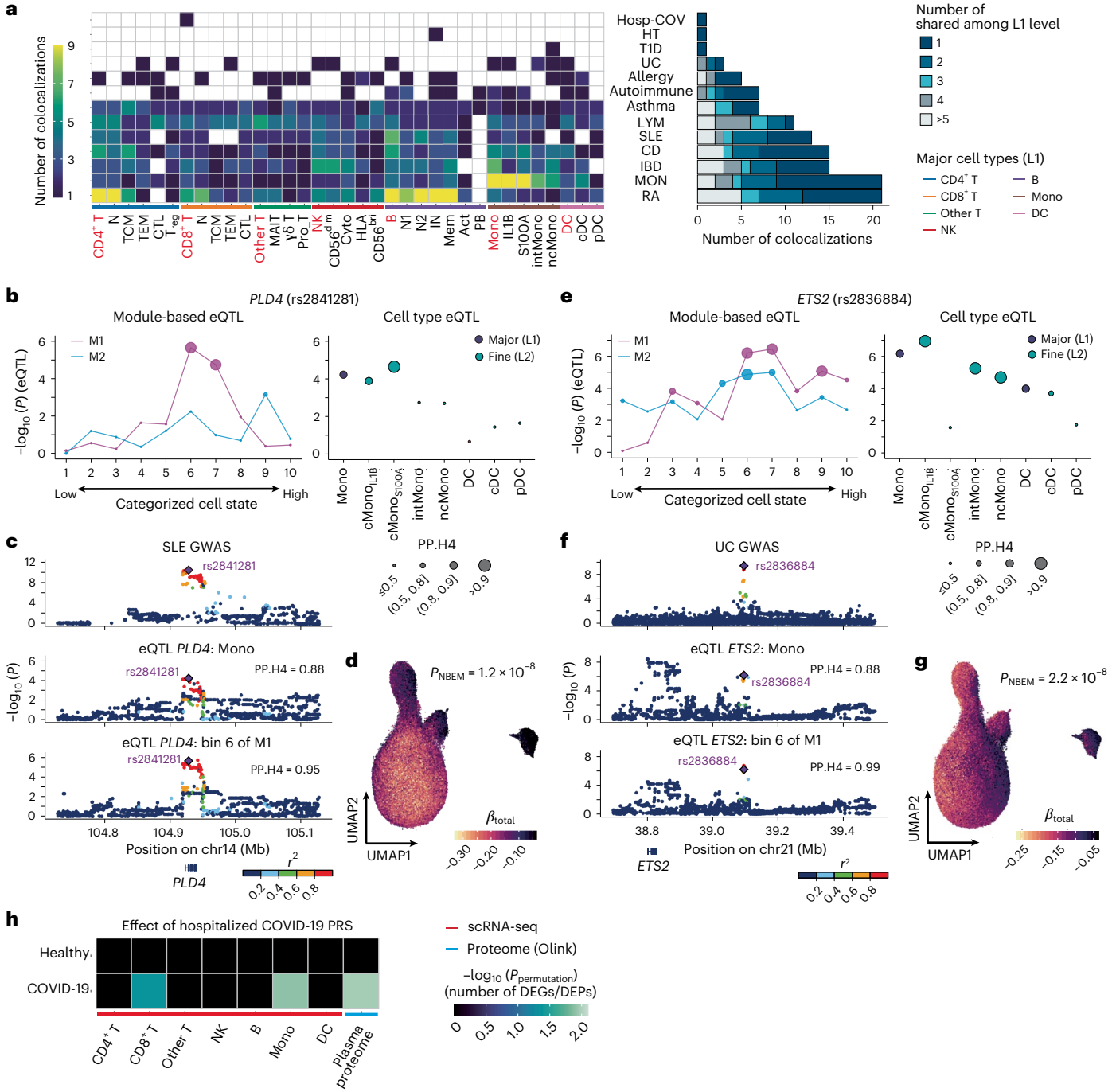


Fig. 5 | Interpretation of GWAS results using OASIS data. a, Number of GWAS loci significantly colocalized with an eQTL for each cell type–trait combination (left). Cell types of the major cell type (L1) are colored red. Number of GWAS loci colocalized with an eQTL and how many times the colocalization was shared among major cell types for each trait (right). GWAS abbreviations: CD, Crohn's disease; Hosp-COV, hospitalized COVID-19; HT, hyperthyroidism; IBD, inflammatory bowel diseases; LYM, lymphocyte count; MON, monocyte count; RA, rheumatoid arthritis; T1D, type 1 diabetes; UC, ulcerative colitis. **b**, Dynamic eQTLs and *cis*-eQTLs per cell type of rs2841281 for *PLD4*. Dot size represents PP.H4 of eQTL–GWAS colocalizations. **c**, Regional association plots of systemic lupus erythematosus GWAS, *cis*-eQTL of monocytes (L1) and dynamic eQTLs in bin 6 of module 1 for the *PLD4* locus. Chr, chromosome. **d**, UMAP represents cell state-dependent eQTL strength (β_{total}) for each cell calculated as a sum of the effect sizes of genotype and genotype × hPCs. Labeled *P* values are derived from NBME analysis. **e**, Dynamic eQTLs and *cis*-eQTLs per cell type of rs2836884 for *ETS2*. **f**, Regional association plots of ulcerative colitis GWAS, *cis*-eQTL of monocytes (L1) and dynamic eQTL in bin 6 of module 1 for the *ETS2* locus. **g**, UMAP represents cell state-dependent eQTL strength (β_{total}) for each cell. Uncorrected *P* values from two-sided tests are shown in **b**–**g**. **h**, The effect of hospitalized COVID-19 PRS on transcriptomic and proteomic levels separately in patients with COVID-19 and healthy individuals. One-sided *P* values were calculated with 1,000 permutations for the number of differentially expressed genes (DEGs) and differentially expressed proteins (DEPs).

https://doi.org/10.1038/s41588-025-02266-3

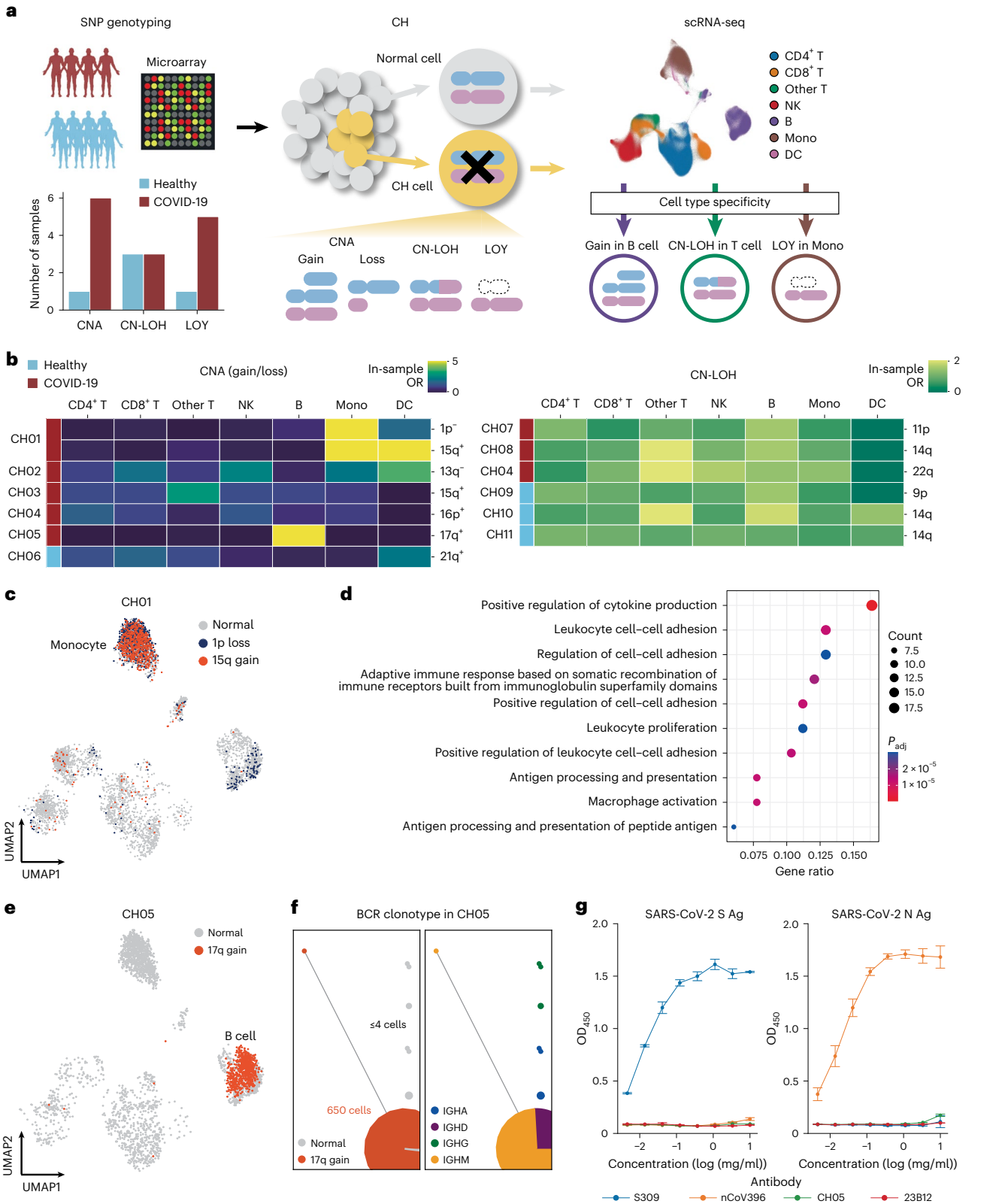


Fig. 6 | Single-cell deconvolution of mCAs. a, Schematic overview of single-cell deconvolution of CH including mCAs and LOY by integrating SNP array and scRNA-seq data. **b**, Heatmaps showing in-sample ORs of each cell type containing cells with CNAs (left) and CN-LOHs (right). **c**, UMAP embedding of CH01 scRNA-seq data colored by three clones. **d**, Top ten enriched biological pathways of upregulated DEGs in monocytes of CH01 with 1p loss. Dot color indicates statistical significance of the enrichment (adjusted *P* values via the Benjamini–Hochberg method), and dot size represents the gene count assigned to each term. **e**, UMAP embedding of CH05 scRNA-seq data colored by two clones. **f**, Network plots showing the similarity of complementarity-determining region 3 (CDR3) amino acid sequences in BCR heavy and light chains of CH05 colored by clone (left) and isotype (right). Clonotype clusters with clonal size >1 are selected. **g**, Reactivity of antibodies against SARS-CoV-2 antigens (Ag) in enzyme-linked immunosorbent assays. Dots denote mean, and error bars show s.d. measured in triplicate. S309, anti-SARS-CoV-2 S immunoglobulin G (IgG)1; CH05, recombinant antibody derived from the CH05 BCR clonotype with 17q gain; nCoV396, anti-SARS-CoV-2 N IgG1; 23B12, anti-*Candida albicans* IgG1; OD_{450}, optical density at 450 nm. Panel **a** created with BioRender.com.

https://doi.org/10.1038/s41588-025-02266-3

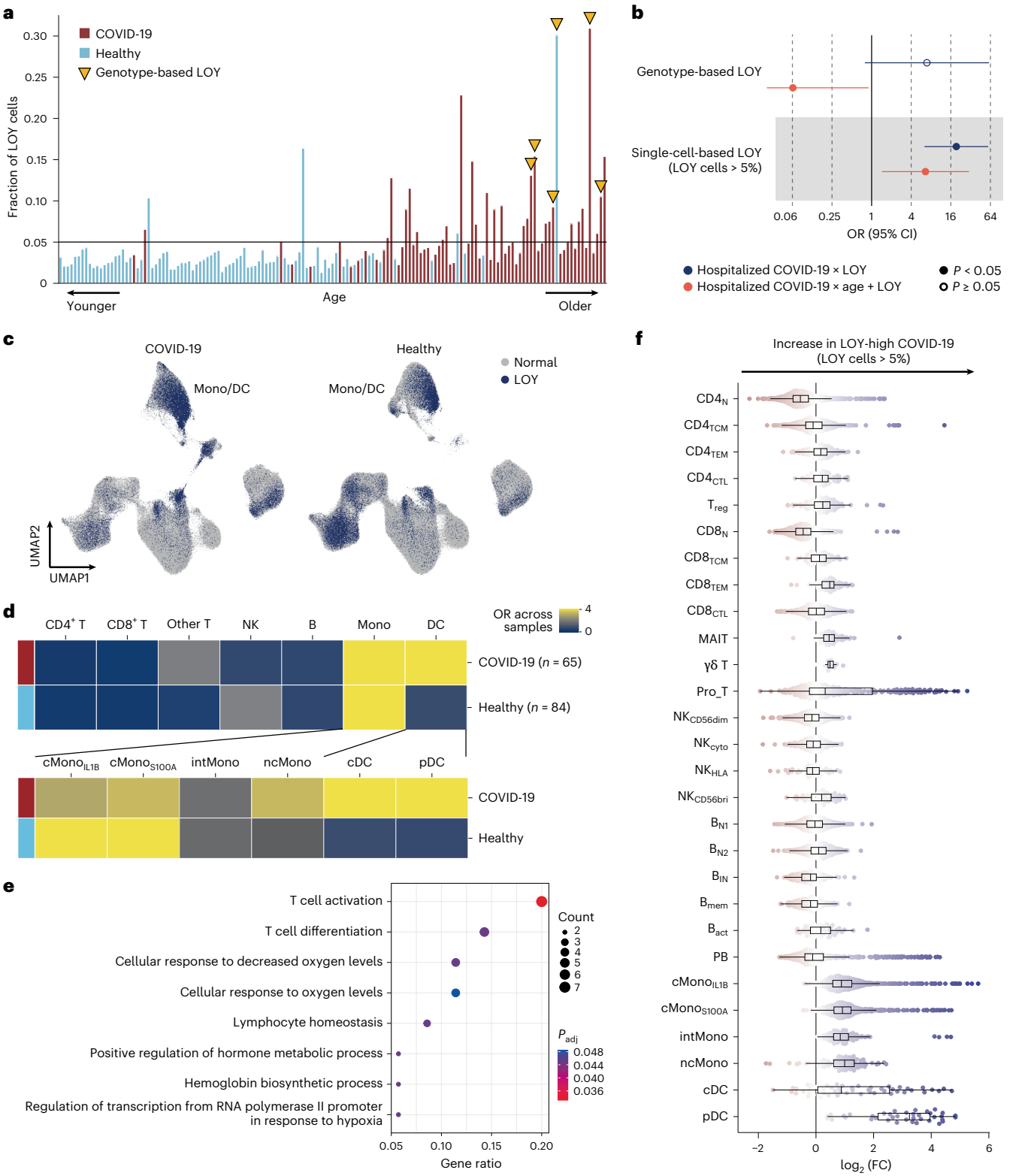


Fig. 7 | Biological effects of mosaic LOY. a, Fractions of cells with LOY in each male sample sorted by age (*x* axis). Yellow triangles denote samples with LOY detected using SNP array data. **b**, Forest plot of associations between LOY and COVID-19 risk. ORs of hospitalized COVID-19 were calculated using univariate logistic regression and multivariate logistic regression to account for the effect of age. Dots represent ORs, and bars represent 95% confidence intervals. We used uncorrected *P* values from two-sided tests. **c**, UMAP embeddings of PBMCs from patients with COVID-19 (left) and healthy individuals (right) colored according to the presence of LOY. **d**, Heatmaps showing ORs across samples of each cell type containing cells with LOY in patients with COVID-19 and healthy individuals. **e**, Enriched biological pathways of upregulated DEGs in cells with LOY compared to normal cells with $cMono_{IL1B}$ cells of patients with COVID-19 and healthy individuals. Dot color indicates statistical significance of the enrichment (adjusted *P* values via the Benjamini–Hochberg method), and dot size represents the gene count assigned to each term. **f**, Beeswarm plot and box plot describing the distribution of adjusted $\log_2$ (fold change (FC)) values between LOY-high and LOY-low samples from patients with COVID-19 in neighborhoods ($n = 12{,}141$) from L2 cell types. Nodes are neighborhoods, colored by their $\log_2$ (FC) adjusted by age and sequencing group. Boxes denote the IQR, medians are shown as vertical bars, and whiskers extend to 1.5 times the IQR.

https://doi.org/10.1038/s41588-025-02266-3

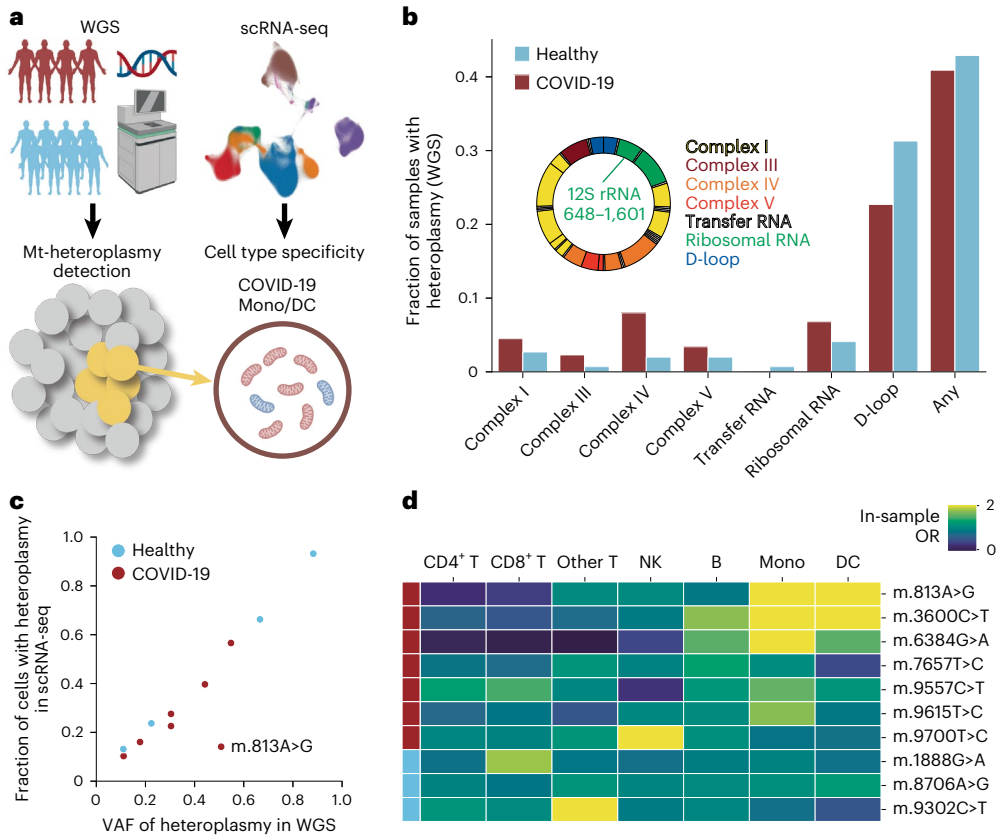


Fig. 8 | The landscape of mitochondrial heteroplasmy. a, Schematic overview of single-cell deconvolution of mt-heteroplasmy by integrating WGS and scRNA-seq data. **b**, Fractions of samples with mt-heteroplasmy detected using WGS data by mtDNA region. Mitochondrial gene positions are indicated in the Circos plot. rRNA, ribosomal RNA. **c**, Correlations between the fractions of cells with heteroplasmy in scRNA-seq data and VAFs of heteroplasmy in WGS data. **d**, Heatmap showing in-sample ORs of each cell type containing cells with mt-heteroplasmy. Panel **a** created with BioRender.com.

Monocyte-specific accumulation of LOY

Next, we evaluated the biological effects of LOY on the immune system. Using scRNA-seq data from male samples ($n = 149$), we defined cells with LOY as cells without expression from the male-specific region of the Y chromosome[15,22], which enabled quantitative estimation of the LOY status of each male. Older men showed larger fractions of cells with LOY. Six male samples with LOY detected using SNP array data (that is, genotype-based binary LOY estimation) contained relatively more cells with LOY (median of 14.2% versus 3.3%; Fig. 7a). The single-cell-based approach was thus more sensitive to detect LOY and provide biological insights. For single-cell-based LOY, all male samples were classified as LOY high (LOY cells > 5%; $n = 36$) or LOY low (LOY cells < 5%; $n = 113$). Using this classification, we performed a regression analysis to answer controversial discussions on whether LOY is associated with COVID-19 risk[22]. While genotype-based LOY did not show an independent impact, single-cell-based LOY was significantly associated with the risk of hospitalized COVID-19 (OR = 6.6, 95% confidence interval = 1.4–30.2 after age adjustment; Fig. 7b).

Cells with LOY were enriched in monocytes across patients with COVID-19 (OR = 4.0) and healthy individuals (OR = 4.5; Fig. 7c,d and Supplementary Table 22). Upregulated DEGs of $cMono_{IL1B}$ cells with LOY were significantly enriched in T cell-related pathways (Fig. 7e and Extended Data Fig. 10a). Next, we performed a differential abundance analysis between LOY-high ($n = 32$) and LOY-low ($n = 33$) samples from patients with COVID-19 using Milo[37]. We found an increase in monocytes and dendritic cells and a decrease in naive T cells for LOY-high samples (Fig. 7f), and a similar trend was observed in healthy individuals (Extended Data Fig. 10b). Comparison of cell proportions between the two groups suggested that a decrease in naive T cells was prominent in $CD4^+$ T cells ($P = 0.013$) and the proportion of regulatory T cells was significantly higher in LOY-high patients with COVID-19 ($P = 7.4 \times 10^{-4}$; Extended Data Fig. 10c). In aggregate, these data suggest that monocytes with LOY might affect the immune response via compositional changes in T cells.

Context-specific enrichment of mitochondrial heteroplasmy

We also investigated the landscape of mt-heteroplasmy in immune cells by integrating our genomics and scRNA-seq data (Fig. 8a). The Genome Analysis Toolkit (GATK)-based pipeline[52] detected 36 (40.9%) patients with COVID-19 and 63 (42.9%) healthy individuals with mt-heteroplasmy (variant allele frequency (VAF) > 0.1) using WGS data (Fig. 8b). Referring to allelic information of the predefined WGS-based heteroplasmy embedded in the scRNA-seq raw reads, we could successfully perform clonal assignment of each cell for seven patients with COVID-19 and four healthy individuals. The proportion of cells with mt-heteroplasmy in scRNA-seq reads was strongly associated with VAFs of mt-heteroplasmy in WGS data except for m.813A>G (Fig. 8c). We evaluated cell type specificity of heteroplasmic cells for these samples, except for one healthy sample in which almost all cells had mt-heteroplasmy. We found that the cells with mt-heteroplasmy were enriched in monocytes and dendritic cells specifically for patients with COVID-19 (Fig. 8d). Of note, the heteroplasmy of m.813A>G, with a large discrepancy between the heteroplasmic cell fraction and VAF, showed stronger cell type specificity (OR = 13.3 for monocytes, OR = 6.8 for dendritic cells; Supplementary Table 23). Given its location in 12S ribosomal RNA, the mutation hotspot for aminoglycoside ototoxicity that is implicated in human disease[53], this heteroplasmy might affect cell proliferation or viability in a cell type-specific manner, causing abnormal mitochondrial function.

This observation highlighted the advantage of our multi-layered omics focusing on both the mutation itself (that is, genomics) and mutated cells (that is, scRNA-seq). Although we did not detect any DEGs to distinguish heteroplasmic cells from normal ones (false discovery rate (FDR) < 0.05), COVID-19-specific enrichment suggested that mt-heteroplasmy may underlie the biological mechanisms of COVID-19.

Discussion

Here, we constructed a single-cell data atlas of >1.5 million PBMCs with multi-layer omics in Japanese individuals. Our integrative analyses of single-cell data revealed that immune cells were dynamically regulated in a cell state-dependent manner defined by multiomics profiles, leading to a better understanding of the pathogenesis of COVID-19 and autoimmune diseases.

We comprehensively mapped germline genetic effects on gene expression in immune cells and observed that eQTLs were dynamically regulated across continuous cell states. We also elucidated HLA and genome-wide regulation of TCR and BCR repertoires in a cell type- and context-specific manner. Thus, including cohorts of diseased and healthy individuals contributes to including cells with a variety of biological conditions. The epigenetic properties of dynamic eQTLs were distinct from those of *cis*-eQTLs, and colocalization analysis highlighted the value of considering genetic regulation of gene expression across continuous cell states to understand GWAS signals. In addition, as in the example of rs11080327 for *SLFN5*, single-cell resolution eQTL modeling allowed us to interpret genetic regulatory mechanisms at finer resolution than bulk and pseudobulk anlaysis[7,33]. Moreover, single-cell technology finely resolved associations between human disease-related gut bacteria and peripheral immune cell abundance.

We deconvoluted a variety of somatic events at single-cell resolution and showed their heterogeneous functional impacts on immune cells, which increased our insights into the underlying mechanisms of COVID-19 severity. As in previous studies of a single gene mutation in hematopoiesis[23] or mCAs in the normal adrenal gland[24], capturing the mutational status of individual cells has enabled us to compare mutant and normal cells within individuals, which is easier when the number of cells per sample is large. Investigating apparently healthy samples or tissues showing one or a few somatic mutations can directly reveal the phenotypic effects of the mutation, and that is not the case for cancer cells with a considerably higher burden of mutations. Of note, some observations of mCAs and mt-heteroplasmy in this study may be a result of the immunologically strong stimulus (that is, SARS-CoV-2 infection), implying the need for a study design that takes context into account.

In conclusion, we demonstrate the importance of interpreting biological phenomena by integrating features from multi-layer omics at single-cell resolution. This resource in an East Asian population will contribute to equity in global diversity of human genomics and help researchers better understand complex human traits.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-025-02266-3.

References

1. Aguet, F. et al. Molecular quantitative trait loci. *Nat. Rev. Methods Primers* **3**, 5 (2023).
2. Aguet, F. et al. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
3. Strober, B. J. et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **18**, 1318–1330 (2020).
4. Sun, B. B. et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
5. Qin, Y. et al. Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort. *Nat. Genet.* **54**, 134–142 (2022).
6. Yazar, S. et al. Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science* **376**, eabf3041 (2022).
7. Perez, R. K. et al. Single-cell RNA-seq reveals cell type-specific molecular and genetic associations to lupus. *Science* **376**, eabf1970 (2022).
8. Soskic, B. et al. Immune disease risk variants regulate gene expression dynamics during CD4$^+$ T cell activation. *Nat. Genet.* **54**, 817–826 (2022).
9. Nathan, A. et al. Single-cell eQTL models reveal dynamic T cell state dependence of disease loci. *Nature* **606**, 120–128 (2022).
10. Kumasaka, N. et al. Mapping interindividual dynamics of innate immune response at single-cell resolution. *Nat. Genet.* **55**, 1066–1075 (2023).
11. Zheng, L. et al. Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science* **374**, abe6474 (2021).
12. Yoshida, M. et al. Local and systemic responses to SARS-CoV-2 infection in children and adults. *Nature* **602**, 321–327 (2022).
13. Oelen, R. et al. Single-cell RNA-sequencing of peripheral blood mononuclear cells reveals widespread, context-specific gene expression regulation upon pathogenic exposure. *Nat. Commun.* **13**, 3267 (2022).
14. Loh, P., Genovese, G. & McCarroll, S. A. Monogenic and polygenic inheritance become instruments for clonal selection. *Nature* **584**, 136–141 (2020).
15. Thompson, D. J. et al. Genetic predisposition to mosaic Y chromosome loss in blood. *Nature* **575**, 652–657 (2019).
16. Laurie, C. C. et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat. Genet.* **44**, 642–650 (2012).
17. Genovese, G. et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**, 2477–2487 (2014).
18. Saiki, R. et al. Combined landscape of single-nucleotide variants and copy number alterations in clonal hematopoiesis. *Nat. Med.* **27**, 1239–1249 (2021).
19. Jaiswal, S. et al. Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. *N. Engl. J. Med.* **377**, 111–121 (2017).
20. Sano, S. et al. Hematopoietic loss of Y chromosome leads to cardiac fibrosis and heart failure mortality. *Science* **377**, 292–297 (2022).
21. Zekavat, S. M. et al. Hematopoietic mosaic chromosomal alterations increase the risk for diverse types of infection. *Nat. Med.* **27**, 1012–1024 (2021).
22. Bruhn-Olszewska, B. et al. Loss of Y in leukocytes as a risk factor for critical COVID-19 in men. *Genome Med.* **14**, 139 (2022).
23. Nam, A. S. et al. Single-cell multi-omics of human clonal hematopoiesis reveals that *DNMT3A* R882 mutations perturb early progenitor states through selective hypomethylation. *Nat. Genet.* **54**, 1514–1526 (2022).
24. Gao, T. et al. A pan-tissue survey of mosaic chromosomal alterations in 948 individuals. *Nat. Genet.* **55**, 1901–1911 (2023).
25. Abdel-Hafiz, H. A. et al. Y chromosome loss in cancer drives growth by evasion of adaptive immunity. *Nature* **619**, 624–631 (2023).
26. Lareau, C. A. et al. Single-cell multi-omics of mitochondrial DNA disorders reveals dynamics of purifying selection across human immune cells. *Nat. Genet.* **55**, 1198–1209 (2023).
27. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).

28. Mulder, K. et al. Cross-tissue single-cell landscape of human monocytes and macrophages in health and disease. *Immunity* **54**, 1883–1900 (2021).
29. Turner, J. S. et al. Human germinal centres engage memory and naive B cells after influenza vaccination. *Nature* **586**, 127–132 (2020).
30. Ren, X. et al. COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas. *Cell* **184**, 1895–1913 (2021).
31. Krämer, B. et al. Early IFN-α signatures and persistent dysfunction are distinguishing features of NK cells in severe COVID-19. *Immunity* **54**, 2650–2669 (2021).
32. Alquicira-Hernandez, J., Sathe, A., Ji, H. P., Nguyen, Q. & Powell, J. E. ScPred: accurate supervised method for cell-type classification from single-cell RNA-seq data. *Genome Biol.* **20**, 264 (2019).
33. Ota, M. et al. Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell* **184**, 3006–3021 (2021).
34. Urbut, S. M., Wang, G., Carbonetto, P. & Stephens, M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat. Genet.* **51**, 187–195 (2019).
35. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–329 (2015).
36. Farh, K. K.-H. et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
37. Dann, E., Henderson, N. C., Teichmann, S. A., Morgan, M. D. & Marioni, J. C. Differential abundance testing on single-cell data using *k*-nearest neighbor graphs. *Nat. Biotechnol.* **40**, 245–253 (2022).
38. Franzosa, E. A. et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat. Microbiol.* **4**, 293–305 (2019).
39. Nii, T. et al. Genomic repertoires linked with pathogenic potency of arthritogenic *Prevotella copri* isolated from the gut of patients with rheumatoid arthritis. *Ann. Rheum. Dis.* **82**, 621–629 (2023).
40. Mills, R. H. et al. Multi-omics analyses of the ulcerative colitis gut microbiome link *Bacteroides vulgatus* proteases with disease severity. *Nat. Microbiol.* **7**, 262–276 (2022).
41. La Gruta, N. L., Gras, S., Daley, S. R., Thomas, P. G. & Rossjohn, J. Understanding the drivers of MHC restriction of T cell receptors. *Nat. Rev. Immunol.* **18**, 467–478 (2018).
42. Nagafuchi, Y. et al. Control of naive and effector CD4 T cell receptor repertoires by rheumatoid-arthritis-risk HLA alleles. *J. Autoimmun.* **133**, 102907 (2022).
43. Ishigaki, K. et al. HLA autoimmune risk alleles restrict the hypervariable region of T cell receptors. *Nat. Genet.* **54**, 393–402 (2022).
44. Edahiro, R. et al. Single-cell analyses and host genetics highlight the role of innate immune cells in COVID-19 severity. *Nat. Genet.* **55**, 753–767 (2023).
45. Ruan, Y. et al. Improving polygenic prediction in ancestrally diverse populations. *Nat. Genet.* **54**, 573–580 (2022).
46. The COVID-19 Host Genetics Initiative. A second update on mapping the human genetic architecture of COVID-19. *Nature* **621**, E7–E26 (2023).
47. Namkoong, H. et al. DOCK2 is involved in the host genetics and biology of severe COVID-19. *Nature* **609**, 754–760 (2022).
48. Wang, Q. S. et al. The whole blood transcriptional regulation landscape in 465 COVID-19 infected samples from Japan COVID-19 Task Force. *Nat. Commun.* **13**, 4830 (2022).
49. Loh, P. R. et al. Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations. *Nature* **559**, 350–355 (2018).
50. Gao, T. et al. Haplotype-aware analysis of somatic copy number variations from single-cell transcriptomes. *Nat. Biotechnol.* **41**, 417–426 (2023).
51. Ahern, D. J. et al. A blood atlas of COVID-19 defines hallmarks of disease severity and specificity. *Cell* **185**, 916–938 (2022).
52. Laricchia, K. M. et al. Mitochondrial DNA variation across 56,434 individuals in gnomAD. *Genome Res.* **32**, 569–582 (2022).
53. Zou, T. et al. Impacts of impaired mitochondrial dynamics in hearing loss: potential therapeutic targets. *Front. Neurosci.* **16**, 998507 (2022).

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025

Ryuya Edahiro[1,2,3,167] ✉, Go Sato[1,3,4,167], Tatsuhiko Naito[1,3], Yuya Shirai[1,2,5], Ryunosuke Saiki[6], Kyuto Sonehara[1,3,7], Yoshihiko Tomofuji[1,3,7], Kenichi Yamamoto[1,8,9], Shinichi Namba[1,3,7], Noah Sasa[1,3,10], Genta Nagao[7,11], Qingbo S. Wang[1,7], Yugo Takahashi[1], Takanori Hasegawa[12], Toshihiro Kishikawa[1,10,13], Ken Suzuki[1,14], Yu-Chen Liu[15], Daisuke Motooka[15,16,17], Ayako Takuwa[15], Hiromu Tanaka[11], Shuhei Azekawa[11], Japan COVID-19 Task Force*, Ho Namkoong[18], Ryuji Koike[19], Akinori Kimura[20], Seiya Imoto[21], Satoru Miyano[12], Takanori Kanai[22], Koichi Fukunaga[11], Mamoru Uemura[4], Takayoshi Morita[2,23], Yasuhiro Kato[2,23], Haruhiko Hirata[2], Yoshito Takeda[2], Yuichiro Doki[4], Hidetoshi Eguchi[4], Daisuke Okuzaki[15,16,17,24,25], Shuhei Sakakibara[26,27], Seishi Ogawa[6,28], Atsushi Kumanogoh[2,17,23,24,25,29,168] ✉ & Yukinori Okada[1,3,5,7,30,168] ✉

[1]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan. [2]Department of Respiratory Medicine and Clinical Immunology, Osaka University Graduate School of Medicine, Suita, Japan. [3]Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [4]Department of Gastroenterological Surgery, Osaka University Graduate School of Medicine, Suita, Japan. [5]Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan. [6]Department of Pathology and Tumor Biology, Kyoto University, Kyoto, Japan. [7]Department of Genome Informatics, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan. [8]Department of Pediatrics, Osaka University Graduate School of Medicine, Suita, Japan. [9]Division of Health Sciences, Osaka University Graduate School of Medicine,

Suita, Japan. [10]Department of Otorhinolaryngology—Head and Neck Surgery, Osaka University Graduate School of Medicine, Suita, Japan. [11]Division of Pulmonary Medicine, Department of Medicine, Keio University School of Medicine, Tokyo, Japan. [12]M&D Data Science Center, Tokyo Medical and Dental University, Tokyo, Japan. [13]Department of Head and Neck Surgery, Aichi Cancer Center Hospital, Nagoya, Japan. [14]Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan. [15]Laboratory of Human Immunology (Single Cell Genomics), WPI-IFReC, Osaka University, Suita, Japan. [16]Genome Information Research Center, Research Institute for Microbial Diseases, Osaka University, Suita, Japan. [17]Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Suita, Japan. [18]Department of Infectious Diseases, Keio University School of Medicine, Tokyo, Japan. [19]Health Science Research and Development Center (HeRD), Tokyo Medical and Dental University, Tokyo, Japan. [20]Institute of Science Tokyo, Tokyo, Japan. [21]Division of Health Medical Intelligence, Human Genome Center, the Institute of Medical Science, the University of Tokyo, Tokyo, Japan. [22]Division of Gastroenterology and Hepatology, Department of Medicine, Keio University School of Medicine, Tokyo, Japan. [23]Department of Immunopathology, WPI-IFReC, Osaka University, Suita, Japan. [24]Center for Infectious Disease Education and Research (CiDER), Osaka University, Suita, Japan. [25]Japan Agency for Medical Research and Development–Core Research for Evolutional Science and Technology (AMED–CREST), Osaka University, Suita, Japan. [26]Laboratory of Systems Immunology, Immunology Frontier Research Center, Osaka University, Suita, Japan. [27]Graduate School of Medical Safety Management, Jikei University of Health Care Sciences, Osaka, Japan. [28]Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto University, Kyoto, Japan. [29]Center for Advanced Modalities and DDS (CAMaD), Osaka University, Suita, Japan. [30]Premium Research Institute for Human Metaverse Medicine (WPI-PRIMe), Osaka University, Suita, Japan. [167]These authors contributed equally: Ryuya Edahiro, Go Sato. [168]These authors jointly supervised this work: Atsushi Kumanogoh, Yukinori Okada. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: r.edahiro@imed3.med.osaka-u.ac.jp; kumanogo@imed3.med.osaka-u.ac.jp; yuki-okada@m.u-tokyo.ac.jp

Japan COVID-19 Task Force

Qingbo S. Wang[1,5], Ryuya Edahiro[1,2], Ho Namkoong[18], Takanori Hasegawa[12], Yuya Shirai[1,2], Kyuto Sonehara[1,3,7], Hiromu Tanaka[11], Ho Lee[11], Ryunosuke Saiki[6], Takayoshi Hyugaji[21], Eigo Shimizu[21], Kotoe Katayama[21], Masahiro Kanai[31], Tatsuhiko Naito[1], Noah Sasa[1,10], Kenichi Yamamoto[1,8,9], Yasuhiro Kato[2,23], Takayoshi Morita[2,23], Kazuhisa Takahashi[32], Norihiro Harada[32], Toshio Naito[33], Makoto Hiki[34,35], Yasushi Matsushita[36], Haruhi Takagi[32], Masako Ichikawa[32], Ai Nakamura[32], Sonoko Harada[32,37], Yuuki Sandhu[32], Hiroki Kabata[11], Katsunori Masaki[11], Hirofumi Kamata[11], Shinnosuke Ikemura[11], Shotaro Chubachi[11], Satoshi Okamori[11], Hideki Terai[11], Atsuho Morita[11], Takanori Asakura[11], Junichi Sasaki[38], Hiroshi Morisaki[39], Yoshifumi Uwamino[40], Kosaku Nanki[22], Sho Uchida[18], Shunsuke Uno[18], Tomoyasu Nishimura[18,41], Takashi Ishiguro[42], Taisuke Isono[42], Shun Shibata[42], Yuma Matsui[42], Chiaki Hosoda[42], Kenji Takano[42], Takashi Nishida[42], Yoichi Kobayashi[42], Yotaro Takaku[42], Noboru Takayanagi[42], Soichiro Ueda[43], Ai Tada[43], Masayoshi Miyawaki[43], Masaomi Yamamoto[43], Eriko Yoshida[43], Reina Hayashi[43], Tomoki Nagasaka[43], Sawako Arai[43], Yutaro Kaneko[43], Kana Sasaki[43], Etsuko Tagaya[44], Masatoshi Kawana[45], Ken Arimura[44], Kunihiko Takahashi[12], Tatsuhiko Anzai[12], Satoshi Ito[12], Akifumi Endo[46], Yuji Uchimura[47], Yasunari Miyazaki[48], Takayuki Honda[48], Tomoya Tateishi[48], Shuji Tohda[49], Naoya Ichimura[49], Kazunari Sonobe[49], Chihiro Tani Sassa[49], Jun Nakajima[49], Yasushi Nakano[50], Yukiko Nakajima[50], Ryusuke Anan[50], Ryosuke Arai[50], Yuko Kurihara[50], Yuko Harada[50], Kazumi Nishio[50], Tetsuya Ueda[50], Masanori Azuma[51], Ryuichi Saito[51], Toshikatsu Sado[51], Yoshimune Miyazaki[51], Ryuichi Sato[51], Yuki Haruta[51], Tadao Nagasaki[51], Yoshinori Yasui[52], Yoshinori Hasegawa[51], Yoshikazu Mutoh[53], Tomoki Kimura[54], Tomonori Sato[54], Reoto Takei[54], Satoshi Hagimoto[54], Yoichiro Noguchi[54], Yasuhiko Yamano[54], Hajime Sasano[54], Sho Ota[54], Yasushi Nakamori[55], Kazuhisa Yoshiya[55], Fukuki Saito[55], Tomoyuki Yoshihara[55], Daiki Wada[55], Hiromu Iwamura[55], Syuji Kanayama[55], Shuhei Maruyama[55], Takashi Yoshiyama[56], Ken Ohta[56], Hiroyuki Kokuto[56], Hideo Ogata[56], Yoshiaki Tanaka[56], Kenichi Arakawa[56], Masafumi Shimoda[56], Takeshi Osawa[56], Hiroki Tateno[57], Isano Hase[57], Shuichi Yoshida[57], Shoji Suzuki[57], Miki Kawada[58], Hirohisa Horinouchi[59], Fumitake Saito[60], Keiko Mitamura[61], Masao Hagihara[62], Junichi Ochi[60], Tomoyuki Uchida[62], Rie Baba[63], Daisuke Arai[63], Takayuki Ogura[63], Hidenori Takahashi[63], Shigehiro Hagiwara[63], Genta Nagao[11], Shunichiro Konishi[63], Ichiro Nakachi[63], Koji Murakami[64], Mitsuhiro Yamada[64], Hisatoshi Sugiura[64], Hirohito Sano[64], Shuichiro Matsumoto[64], Nozomu Kimura[64], Yoshinao Ono[64], Hiroaki Baba[65], Yusuke Suzuki[66], Sohei Nakayama[66], Keita Masuzawa[66], Shinichi Namba[1], Takayuki Shiroyama[2], Yoshimi Noda[2], Takayuki Niitsu[2], Yuichi Adachi[2], Takatoshi Enomoto[2], Saori Amiya[2], Reina Hara[2], Yuta Yamaguchi[2,23], Teruaki Murakami[2,23], Tomoki Kuge[2], Kinnosuke Matsumoto[2], Yuji Yamamoto[2], Makoto Yamamoto[2], Midori Yoneda[2], Kazunori Tomono[67], Kazuto Kato[68], Haruhiko Hirata[2], Yoshito Takeda[2], Hidefumi Koh[69], Tadashi Manabe[69], Yohei Funatsu[69], Fumimaro Ito[69], Takahiro Fukui[69], Keisuke Shinozuka[69], Sumiko Kohashi[69], Masatoshi Miyazaki[69], Tomohisa Shoko[70], Mitsuaki Kojima[70], Tomohiro Adachi[70], Motonao Ishikawa[71], Kenichiro Takahashi[72], Takashi Inoue[73], Toshiyuki Hirano[73], Keigo Kobayashi[73], Hatsuyo Takaoka[73], Kazuyoshi Watanabe[74], Naoki Miyazawa[75], Yasuhiro Kimura[75], Reiko Sado[75], Hideyasu Sugimoto[75], Akane Kamiya[76], Naota Kuwahara[77], Akiko Fujiwara[77], Tomohiro Matsunaga[77], Yoko Sato[77], Takenori Okada[77], Yoshihiro Hirai[78], Hidetoshi Kawashima[78], Atsuya Narita[78], Kazuki Niwa[79], Yoshiyuki Sekikawa[79], Koichi Nishi[80], Masaru Nishitsuji[80], Mayuko Tani[80], Junya Suzuki[80], Hiroki Nakatsumi[80], Takashi Ogura[81], Hideya Kitamura[81], Eri Hagiwara[81], Kota Murohashi[81], Hiroko Okabayashi[81], Takao Mochimaru[82,83], Shigenari Nukaga[82], Ryosuke Satomi[82], Yoshitaka Oyamada[82,83], Nobuaki Mori[84], Tomoya Baba[85], Yasutaka Fukui[85], Mitsuru Odate[85], Shuko Mashimo[85], Yasushi Makino[85], Kazuma Yagi[86], Mizuha Hashiguchi[86], Junko Kagyo[86],

Tetsuya Shiomi[86], Satoshi Fuke[87], Hiroshi Saito[87], Tomoya Tsuchida[88], Shigeki Fujitani[89], Mumon Takita[89], Daiki Morikawa[89], Toru Yoshida[89], Takehiro Izumo[90], Minoru Inomata[90], Naoyuki Kuse[90], Nobuyasu Awano[90], Mari Tone[90], Akihiro Ito[91], Yoshihiko Nakamura[92], Kota Hoshino[92], Junichi Maruyama[92], Hiroyasu Ishikura[92], Tohru Takata[93], Toshio Odani[94], Masaru Amishima[95], Takeshi Hattori[95], Yasuo Shichinohe[96], Takashi Kagaya[97], Toshiyuki Kita[97], Kazuhide Ohta[97], Satoru Sakagami[97], Kiyoshi Koshida[97], Kentaro Hayashi[98], Tetsuo Shimizu[98], Yutaka Kozu[98], Hisato Hiranuma[98], Yasuhiro Gon[98], Namiki Izumi[99], Kaoru Nagata[99], Ken Ueda[99], Reiko Taki[99], Satoko Hanada[99], Kodai Kawamura[100], Kazuya Ichikado[100], Kenta Nishiyama[100], Hiroyuki Muranaka[100], Kazunori Nakamura[100], Naozumi Hashimoto[101], Keiko Wakahara[101], Sakamoto Koji[101], Norihito Omote[101], Akira Ando[101], Nobuhiro Kodama[102], Yasunari Kaneyama[102], Shunsuke Maeda[102], Takashige Kuraki[103], Takemasa Matsumoto[103], Koutaro Yokote[104], Taka-Aki Nakada[105], Ryuzo Abe[105], Taku Oshima[105], Tadanaga Shimada[105], Masahiro Harada[106], Takeshi Takahashi[106], Hiroshi Ono[106], Toshihiro Sakurai[106], Takayuki Shibusawa[106], Yoshifumi Kimizuka[107], Akihiko Kawana[107], Tomoya Sano[107], Chie Watanabe[107], Ryohei Suematsu[107], Hisako Sageshima[108], Ayumi Yoshifuji[109], Kazuto Ito[109], Saeko Takahashi[110], Kota Ishioka[110], Morio Nakamura[110], Makoto Masuda[111], Aya Wakabayashi[111], Hiroki Watanabe[111], Suguru Ueda[111], Masanori Nishikawa[111], Yusuke Chihara[112], Mayumi Takeuchi[112], Keisuke Onoi[112], Jun Shinozuka[112], Atsushi Sueyoshi[112], Yoji Nagasaki[113], Masaki Okamoto[114,115], Sayoko Ishihara[116], Masatoshi Shimo[116], Yoshihisa Tokunaga[114,115], Yu Kusaka[117], Takehiko Ohba[117], Susumu Isogai[117], Aki Ogawa[117], Takuya Inoue[117], Satoru Fukuyama[118], Yoshihiro Eriguchi[119], Akiko Yonekawa[119], Keiko Kan-o[118], Koichiro Matsumoto[118], Kensuke Kanaoka[120], Shoichi Ihara[120], Kiyoshi Komuta[120], Yoshiaki Inoue[121], Shigeru Chiba[122], Kunihiro Yamagata[123], Yuji Hiramatsu[124], Hirayasu Kai[123], Koichiro Asano[125], Tsuyoshi Oguma[125], Yoko Ito[125], Satoru Hashimoto[126], Masaki Yamasaki[126], Yu Kasamatsu[127], Yuko Komase[128], Naoya Hida[128], Takahiro Tsuburai[128], Baku Oyama[128], Minoru Takada[129], Hidenori Kanda[129], Yuichiro Kitagawa[130], Tetsuya Fukuta[130], Takahito Miyake[130], Shozo Yoshida[130], Shinji Ogura[130], Shinji Abe[131], Yuta Kono[131], Yuki Togashi[131], Hiroyuki Takoi[131], Ryota Kikuchi[131], Shinichi Ogawa[132], Tomouki Ogata[132], Shoichiro Ishihara[132], Arihiko Kanehiro[133,134], Shinji Ozaki[133], Yasuko Fuchimoto[133], Sae Wada[133], Nobukazu Fujimoto[133], Kei Nishiyama[135], Mariko Terashima[136], Satoru Beppu[136], Kosuke Yoshida[136], Osamu Narumoto[137], Hideaki Nagai[137], Nobuharu Ooshima[137], Mitsuru Motegi[138], Akira Umeda[139], Kazuya Miyagawa[140], Hisato Shimada[141], Mayu Endo[142], Yoshiyuki Ohira[139], Masafumi Watanabe[143], Sumito Inoue[143], Akira Igarashi[143], Masamichi Sato[143], Hironori Sagara[144], Akihiko Tanaka[144], Shin Ohta[144], Tomoyuki Kimura[144], Yoko Shibata[145], Yoshinori Tanino[145], Takefumi Nikaido[145], Hiroyuki Minemura[145], Yuki Sato[145], Yuichiro Yamada[146], Takuya Hashino[146], Masato Shinoki[146], Hajime Iwagoe[147], Hiroshi Takahashi[148], Kazuhiko Fujii[148], Hiroto Kishi[148], Masayuki Kanai[149], Tomonori Imamura[149], Tatsuya Yamashita[149], Masakiyo Yatomi[150], Toshitaka Maeno[150], Shinichi Hayashi[151], Mai Takahashi[151], Mizuki Kuramochi[151], Isamu Kamimaki[151], Yoshiteru Tominaga[151], Tomoo Ishii[152], Mitsuyoshi Utsugi[153], Akihiro Ono[153], Toru Tanaka[154], Takeru Kashiwada[154], Kazue Fujita[154], Yoshinobu Saito[154], Masahiro Seike[154], Hiroko Watanabe[155], Hiroto Matsuse[156], Norio Kodaka[156], Chihiro Nakano[156], Takeshi Oshio[156], Takatomo Hirouchi[156], Shohei Makino[157], Moritoki Egi[157], Yosuke Omae[158], Yasuhito Nannya[6], Takafumi Ueno[159], Tomomi Takano[160], Kazuhiko Katayama[161], Masumi Ai[162], Atsushi Kumanogoh[2,17,23,24,25], Toshiro Sato[163], Naoki Hasegawa[18], Katsushi Tokunaga[158], Makoto Ishii[11], Ryuji Koike[19], Yuko Kitagawa[164], Akinori Kimura[20], Seiya Imoto[21], Satoru Miyano[12], Seishi Ogawa[6,28,165], Takanori Kanai[22,166], Koichi Fukunaga[11] & Yukinori Okada[1,3,5,7]

[31]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. [32]Department of Respiratory Medicine, Juntendo University Faculty of Medicine and Graduate School of Medicine, Tokyo, Japan. [33]Department of General Medicine, Juntendo University Faculty of Medicine and Graduate School of Medicine, Tokyo, Japan. [34]Department of Emergency and Disaster Medicine, Juntendo University Faculty of Medicine and Graduate School of Medicine, Tokyo, Japan. [35]Department of Cardiovascular Biology and Medicine, Juntendo University Faculty of Medicine and Graduate School of Medicine, Tokyo, Japan. [36]Department of Internal Medicine and Rheumatology, Juntendo University Faculty of Medicine and Graduate School of Medicine, Tokyo, Japan. [37]Atopy (Allergy) Research Center, Juntendo University Graduate School of Medicine, Tokyo, Japan. [38]Department of Emergency and Critical Care Medicine, Keio University School of Medicine, Tokyo, Japan. [39]Department of Anesthesiology, Keio University School of Medicine, Tokyo, Japan. [40]Department of Laboratory Medicine, Keio University School of Medicine, Tokyo, Japan. [41]Keio University Health Center, Tokyo, Japan. [42]Department of Respiratory Medicine, Saitama Cardiovascular and Respiratory Center, Kumagaya, Japan. [43]Internal Medicine, Japan Community Health Care Organization, Saitama Medical Center, Saitama, Japan. [44]Department of Respiratory Medicine, Tokyo Women's Medical University, Tokyo, Japan. [45]Department of General Medicine, Tokyo Women's Medical University, Tokyo, Japan. [46]Clinical Research Center, Tokyo Medical and Dental University Hospital of Medicine, Tokyo, Japan. [47]Department of Medical Informatics, Tokyo Medical and Dental University Hospital of Medicine, Tokyo, Japan. [48]Respiratory Medicine, Tokyo Medical and Dental University, Tokyo, Japan. [49]Clinical Laboratory, Tokyo Medical and Dental University Hospital of Medicine, Tokyo, Japan. [50]Department of Internal Medicine, Kawasaki Municipal Ida Hospital, Kawasaki, Japan. [51]Department of Respiratory Medicine, Osaka Saiseikai Nakatsu Hospital, Osaka, Japan. [52]Department of Infection Control, Osaka Saiseikai Nakatsu Hospital, Osaka, Japan. [53]Department of Infectious Diseases, Tosei General Hospital, Seto, Japan. [54]Department of Respiratory, Allergic Diseases Internal Medicine, Tosei General Hospital, Seto, Japan. [55]Department of Emergency and Critical Care Medicine, Kansai Medical University General Medical Center, Moriguchi, Japan. [56]Japan Anti-Tuberculosis Association, Fukujuji Hospital, Kiyose, Japan. [57]Department of Pulmonary Medicine, Saitama City Hospital, Saitama, Japan. [58]Department of Infectious Diseases, Saitama City Hospital, Saitama, Japan. [59]Department of General Thoracic Surgery, Saitama City Hospital, Saitama, Japan. [60]Department of Pulmonary Medicine, Eiju General Hospital, Tokyo, Japan. [61]Division of Infection Control, Eiju General Hospital, Tokyo, Japan. [62]Department of Hematology, Eiju General Hospital, Tokyo, Japan. [63]Saiseikai Utsunomiya Hospital, Utsunomiya, Japan. [64]Department of Respiratory Medicine, Tohoku University Graduate School of Medicine, Sendai, Japan. [65]Department of Infectious Diseases, Tohoku University Graduate School of Medicine, Sendai, Japan. [66]Department of Respiratory Medicine, Kitasato

University Kitasato Institute Hospital, Tokyo, Japan. [67]Division of Infection Control and Prevention, Osaka University Hospital, Suita, Japan. [68]Department of Biomedical Ethics and Public Policy, Osaka University Graduate School of Medicine, Suita, Japan. [69]Tachikawa Hospital, Tachikawa, Japan. [70]Department of Emergency and Critical Care Medicine, Tokyo Women's Medical University Medical Center East, Tokyo, Japan. [71]Department of Medicine, Tokyo Women's Medical University Medical Center East, Tokyo, Japan. [72]Department of Pediatrics, Tokyo Women's Medical University Medical Center East, Tokyo, Japan. [73]Internal Medicine, Sano Kosei General Hospital, Sano, Japan. [74]Japan Community Health Care Organization, Kanazawa Hospital, Kanazawa, Japan. [75]Department of Respiratory Medicine, Saiseikai Yokohamashi Nanbu Hospital, Yokohama, Japan. [76]Department of Clinical Laboratory, Saiseikai Yokohamashi Nanbu Hospital, Yokohama, Japan. [77]Internal Medicine, Internal Medicine Center, Showa University Koto Toyosu Hospital, Tokyo, Japan. [78]Department of Respiratory Medicine, Japan Organization of Occupational Health and Safety, Kanto Rosai Hospital, Kawasaki, Japan. [79]Department of General Internal Medicine, Japan Organization of Occupational Health and Safety, Kanto Rosai Hospital, Kawasaki, Japan. [80]Ishikawa Prefectural Central Hospital, Kanazawa, Japan. [81]Kanagawa Cardiovascular and Respiratory Center, Yokohama, Japan. [82]Department of Respiratory Medicine, National Hospital Organization Tokyo Medical Center, Tokyo, Japan. [83]Department of Allergy, National Hospital Organization Tokyo Medical Center, Tokyo, Japan. [84]Department of General Internal Medicine and Infectious Diseases, National Hospital Organization Tokyo Medical Center, Tokyo, Japan. [85]Department of Respiratory Medicine, Toyohashi Municipal Hospital, Toyohashi, Japan. [86]Keiyu Hospital, Yokohama, Japan. [87]Department of Respiratory Medicine, KKR Sapporo Medical Center, Sapporo, Japan. [88]Division of General Internal Medicine, Department of Internal Medicine, St. Marianna University School of Medicine, Kawasaki, Japan. [89]Department of Emergency and Critical Care Medicine, St. Marianna University School of Medicine, Kawasaki, Japan. [90]Japanese Red Cross Medical Center, Tokyo, Japan. [91]Matsumoto City Hospital, Matsumoto, Japan. [92]Department of Emergency and Critical Care Medicine, Faculty of Medicine, Fukuoka University, Fukuoka, Japan. [93]Department of Infection Control, Fukuoka University Hospital, Fukuoka, Japan. [94]Department of Rheumatology, National Hospital Organization Hokkaido Medical Center, Sapporo, Japan. [95]Department of Respiratory Medicine, National Hospital Organization Hokkaido Medical Center, Sapporo, Japan. [96]Department of Emergency and Critical Care Medicine, National Hospital Organization Hokkaido Medical Center, Sapporo, Japan. [97]National Hospital Organization Kanazawa Medical Center, Kanazawa, Japan. [98]Division of Respiratory Medicine, Department of Internal Medicine, Nihon University School of Medicine, Tokyo, Japan. [99]Musashino Red Cross Hospital, Musashino, Japan. [100]Division of Respiratory Medicine, Social Welfare Organization Saiseikai Imperial Gift Foundation, Inc., Saiseikai Kumamoto Hospital, Kumamoto, Japan. [101]Department of Respiratory Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan. [102]Department of Internal Medicine, Fukuoka Tokushukai Hospital, Kasuga, Japan. [103]Respiratory Medicine, Fukuoka Tokushukai Hospital, Kasuga, Japan. [104]Department of Endocrinology, Hematology and Gerontology, Chiba University Graduate School of Medicine, Chiba, Japan. [105]Department of Emergency and Critical Care Medicine, Chiba University Graduate School of Medicine, Chiba, Japan. [106]National Hospital Organization Kumamoto Medical Center, Kumamoto, Japan. [107]Division of Infectious Diseases and Respiratory Medicine, Department of Internal Medicine, National Defense Medical College, Tokorozawa, Japan. [108]Sapporo City General Hospital, Sapporo, Japan. [109]Department of Internal Medicine, Tokyo Saiseikai Central Hospital, Tokyo, Japan. [110]Department of Pulmonary Medicine, Tokyo Saiseikai Central Hospital, Tokyo, Japan. [111]Department of Respiratory Medicine, Fujisawa City Hospital, Fujisawa, Japan. [112]Uji-Tokushukai Medical Center, Uji, Japan. [113]Department of Infectious Disease and Clinical Research Institute, National Hospital Organization Kyushu Medical Center, Fukuoka, Japan. [114]Department of Respirology, National Hospital Organization Kyushu Medical Center, Fukuoka, Japan. [115]Division of Respirology, Rheumatology and Neurology, Department of Internal Medicine, Kurume University School of Medicine, Kurume, Japan. [116]Department of Infectious Disease, National Hospital Organization Kyushu Medical Center, Fukuoka, Japan. [117]Ome Municipal General Hospital, Ome, Japan. [118]Research Institute for Diseases of the Chest, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. [119]Department of Medicine and Biosystemic Science, Kyushu University Graduate School of Medical Sciences, Fukuoka, Japan. [120]Daini Osaka Police Hospital, Osaka, Japan. [121]Department of Emergency and Critical Care Medicine, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan. [122]Department of Hematology, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan. [123]Department of Nephrology, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan. [124]Department of Cardiovascular Surgery, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan. [125]Division of Pulmonary Medicine, Department of Medicine, Tokai University School of Medicine, Isehara, Japan. [126]Department of Anesthesiology and Intensive Care Medicine, Kyoto Prefectural University of Medicine, Kyoto, Japan. [127]Department of Infection Control and Laboratory Medicine, Kyoto Prefectural University of Medicine, Kyoto, Japan. [128]Department of Respiratory Internal Medicine, St. Marianna University School of Medicine, Yokohama City Seibu Hospital, Yokohama, Japan. [129]Kinshukai Hanwa the Second Hospital, Osaka, Japan. [130]Emergency and Disaster Medicine, Gifu University School of Medicine Graduate School of Medicine, Gifu, Japan. [131]Department of Respiratory Medicine, Tokyo Medical University Hospital, Tokyo, Japan. [132]JA Toride medical hospital, Toride, Japan. [133]Okayama Rosai Hospital, Okayama, Japan. [134]Himeji St. Mary's Hospital, Himeji, Japan. [135]Emergency and Critical Care, Niigata University, Niigata, Japan. [136]Emergency and Critical Care Center, National Hospital Organization Kyoto Medical Center, Kyoto, Japan. [137]National Hospital Organization Tokyo Hospital Hospital, Kiyose, Japan. [138]Fujioka General Hospital, Fujioka, Japan. [139]Department of General Medicine, School of Medicine, International University of Health and Welfare Shioya Hospital, Ohtawara, Japan. [140]Department of Pharmacology, School of Pharmacy, International University of Health and Welfare Shioya Hospital, Ohtawara, Japan. [141]Department of Respiratory Medicine, International University of Health and Welfare Shioya Hospital, Ohtawara, Japan. [142]Department of Clinical Laboratory, International University of Health and Welfare Shioya Hospital, Ohtawara, Japan. [143]Department of Cardiology, Pulmonology and Nephrology, Yamagata University Faculty of Medicine, Yamagata, Japan. [144]Division of Respiratory Medicine and Allergology, Department of Medicine, School of Medicine, Showa University, Tokyo, Japan. [145]Department of Pulmonary Medicine, Fukushima Medical University, Fukushima, Japan. [146]Kansai Electric Power Hospital, Osaka, Japan. [147]Division of Infectious Diseases, Kumamoto City Hospital, Kumamoto, Japan. [148]Department of Respiratory Medicine, Kumamoto City Hospital, Kumamoto, Japan. [149]Department of Emergency and Critical Care Medicine, Tokyo Metropolitan Police Hospital, Tokyo, Japan. [150]Department of Respiratory Medicine, Gunma University Graduate School of Medicine, Maebashi, Japan. [151]National Hospital Organization Saitama Hospital, Wako, Japan. [152]Tokyo Medical University Ibaraki Medical Center, Inashiki, Japan. [153]Department of Internal Medicine, Kiryu Kosei General Hospital, Kiryu, Japan. [154]Department of Pulmonary Medicine and Oncology, Graduate School of Medicine, Nippon Medical School, Tokyo, Japan. [155]Division of Respiratory Medicine, Tsukuba Kinen General Hospital, Tsukuba, Japan. [156]Division of Respiratory Medicine, Department of Internal Medicine, Toho University Ohashi Medical Center, Tokyo, Japan. [157]Division of Anesthesiology, Department of Surgery Related, Kobe University Graduate School of Medicine, Kobe, Japan. [158]Genome Medical Science Project (Toyama), National Center for Global Health and Medicine, Tokyo, Japan. [159]Department of Biomolecular Engineering, Graduate School of Tokyo Institute of Technology, Tokyo, Japan. [160]Laboratory of Veterinary Infectious Disease, School of Veterinary Medicine, Kitasato University, Aomori, Japan. [161]Laboratory of Viral Infection, Department of Infection Control and Immunology, Ōmura Satoshi Memorial Institute and Graduate School of Infection Control Sciences, Kitasato University, Tokyo, Japan. [162]Department of Insured Medical Care Management, Tokyo Medical and Dental University Hospital of Medicine, Tokyo, Japan. [163]Department of Organoid Medicine, Keio University School of Medicine, Tokyo, Japan. [164]Department of Surgery, Keio University School of Medicine, Tokyo, Japan. [165]Department of Medicine, Center for Hematology and Regenerative Medicine, Karolinska Institute, Stockholm, Sweden. [166]AMED–CREST, Tokyo, Japan.

Methods

Ethics and specimen collection of OASIS samples

Peripheral blood samples were obtained from patients with COVID-19 (n = 88) and healthy controls (n = 147) at Osaka University Hospital. Patients with COVID-19 were further categorized into moderate (n = 9) and severe (n = 79) groups according to disease severity based on the highest score on the World Health Organization Ordinal Scale for Clinical Improvement ever present[54]. Detailed clinical data are summarized in Supplementary Table 1. Some of the participants ($n_{COVID\text{-}19}$ = 73, $n_{control}$ = 75) are described elsewhere[44,47]. There was one pair of blood relatives among the healthy participants; therefore one of them was excluded from the eQTL analysis. One patient with COVID-19 had a karyotype abnormality and was excluded from the X chromosome eQTL analysis and differential gene and protein expression analyses. This study strictly follows the principles of the Declaration of Helsinki, with written informed consent obtained from all participants before sample collection according to regular principles. Ethical approvals were obtained from the Ethics Committee of Osaka University (approval no. 734). There was no compensation for participants.

Sequencing, alignment, quantification and quality control of scRNA-seq

Single-cell suspensions were processed through the 10x Genomics Chromium Controller. Droplet libraries were processed using Cell Ranger 5.0.0 (10x Genomics). Sequencing reads were aligned with STAR (version 2.7.2a)[55] using the GRCh38 human reference genome. Please see the Supplementary Methods for details.

Integration and manual annotation of scRNA-seq

Details are described in the Supplementary Methods.

Whole-genome sequencing data processing

DNA samples isolated from whole blood were sequenced at Macrogen Japan. DNA quantity was measured by PicoGreen, and DNA degradation was assessed by gel electrophoresis. All libraries were constructed using the TruSeq DNA PCR-Free Library Preparation Kit according to the manufacturer's protocols. Libraries were sequenced on the HiSeq X or NovaSeq 6000 system (Illumina), producing paired-end reads 2 × 150 bp in length at an average depth of 19.4×. Sequenced reads were aligned against the Genome Reference Consortium human genome build 38 using BWA-MEM with the ALT-aware mode (version 0.7.17). For more details, please see the Supplementary Methods.

SNP array genotyping

Genomic DNA was genotyped with the use of the Infinium Asian Screening Array (Illumina). This genotyping array was built using an East Asian reference panel including whole-genome sequences, designed for effectively capturing genetic variation in East Asian populations. To increase the sample size used for subsequent statistical haplotype phasing, the accuracy of which determines the sensitivity for mCA detection using MoChA (18 May 2022)[14,49], we merged the publicly available genotyping data of 54,405 Japanese individuals generated by the BioBank Japan Project using the same genotyping array. For more details, please see the Supplementary Methods.

Mapping of single-cell *cis*-eQTLs with the pseudobulk approach

First, we performed single-cell-level normalization using scran (version 1.18.5)[56] separately for each major cell type (L1). We only kept genes with nonzero expression (unique molecular identifier (UMI) count > 0) in more than 1% of cells in each major cell type. In the analysis of subclusters (L2) within each major cell type, the same genes were retained. Gene expression per sample per cell type (that is, pseudobulk count) was calculated as the mean of $\log_2$-transformed normalized expression across cells and then normalized across samples using an inverse normal transform in each cell type. Samples with more than ten cells in a cell type were considered in the analysis of the corresponding cell type (Supplementary Fig. 2b).

To identify *cis*-eQTLs, we used tensorQTL (version 1.0.7)[57] to run a linear regression for each SNP–gene pair. We restricted our search to variants within 1 Mb of the transcription start site of each gene and with MAF > 0.05. The top 15 gene expression PCs, the top two genetic PCs, age, sex, 10x chemistry (version 1 or 2) and clinical status (COVID-19 or healthy) were used as covariates for eQTL analysis. To correct for the number of association tests performed per gene, we used a *cis* permutation pass per gene with 1,000 permutations. Finally, to correct for the number of genes tested and identify significant eGenes, we performed a genome-wide q-value correction for the top associated SNP–gene pair, setting a q-value threshold of 0.05. We subsequently mapped conditionally independent *cis*-QTLs using the stepwise regression procedure with tensorQTL. Details on downsampling sample sizes and cell counts per sample in eQTL mapping are described in the Supplementary Methods.

Generation of genotypes and single-cell eQTL mapping in OneK1K

Details are described in the Supplementary Methods.

Comparison with *cis*-eQTLs of OneK1K

We assessed eQTL overlaps and sharing between our cohort and OneK1K[6] using the multivariate adaptive shrinkage method[34]. For more details, please see the Supplementary Methods.

Epigenetic marker enrichment analysis for *cis*-eQTLs

To reveal the functional characterization of *cis*-eQTLs, we evaluated the enrichment of significant eQTLs to epigenome marks (that is, Roadmap annotation)[35]. The chromatin state data were obtained from the Roadmap Epigenomics project. We used 18-state models. We considered TssA, TssFlnk, TssFlnkU and TssFlnkD as promoter regions and EnhA1, EnhA2, EnhG1, EnhG2 and EnhWk as enhancer regions. For analysis of enhancer or promoter enrichment, we used eight immune cell subsets from peripheral blood in Roadmap project data (that is, primary monocytes (Roadmap ID E029), primary B cells (E032), primary helper T memory cells (E037), primary naive helper T cells (E038), primary regulatory T cells (E044), primary natural killer cells (E046), primary $CD8^+$ naive T cells (E047) and primary $CD8^+$ T memory cells (E048)). To assess the enrichment of primary *cis*-eQTLs in epigenome marks, we compared the overlap with epigenome marks between primary significant *cis*-eQTLs of each cell type and all the variants used in the analysis for that cell type. For more details, please see the Supplementary Methods.

Metagenomics analysis

Phenol–chloroform DNA extraction and subsequent metagenome shotgun sequencing were newly performed (dataset 3) or performed in previous studies (datasets 1 and 2)[58,59]. Details on metagenomics analysis are described in the Supplementary Methods.

TCR and BCR repertoire analysis

Droplet-based sequencing data for TCR sequences and BCR sequences were aligned and quantified using 5.0.0 (10x Genomics) against the GRCh38 human VDJ reference genome. Filtered annotated contigs for TCR sequences and BCR sequences were analyzed using Scirpy (version 0.10.0)[60]. For more details, please see the Supplementary Methods.

HLA and genome-wide association analysis with TCR and BCR repertoires

For HLA association analysis, we performed HLA imputation for genotyped SNPs in the MHC region with the HLA reference panel of the Japanese population (n = 1,118) constructed in a previous study[61] using

DEEP*HLA[62,63]. We targeted the amino acid variants of class I and II HLA genes (class I, *HLA-A*, *HLA-C* and *HLA-B*; class II, *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1* and *HLA-DPB1*). Variants imputed with *an* R^2 imputation quality score in tenfold cross-validation >0.7 and MAF > 0.05 were used for the analyses.

We explored the association between HLA variants and TRAV and TRBV gene usage in each cell type. We only used productive clones for which all the TCR genes were identified. We calculated the usage frequency of each TCR V gene in each cell type by summing up clones that were derived from that gene. We summed up clones regardless of the clone size, unless otherwise specified, to sensitively capture the central restriction between HLA and TCR[43]. We only targeted TCR V genes that were observed in more than 80% of samples in cell types of interest.

We investigated the association between HLA variants and TCR V gene usage based on the Wald test for a negative binomial regression model using DESeq2 (version 1.30.1)[64] for each cell type. We included the same covariates as those used in the *cis*-eQTL analysis, excluding PCs for gene expression, in the model. For the HLA association analysis, we empirically calculated the study-wide significance threshold using a permutation procedure[65] and obtained study-wide significance thresholds of $P = 1.1 \times 10^{-5}$ and 1.2×10^{-5} for TRAV and TRBV genes, respectively. We also investigated the change in TCR V genes according to each variant under the condition of COVID-19 by adding an interaction term between COVID-19 status and each variant into the model. In this interaction analysis, we used the usage of TCR V genes weighted by their clone sizes to consider the effect of clonal expansion. We performed these analyses only for entire (that is, $CD4^+$ T at L1), naive ($CD4_N$) and $CD4_{TCM}$ and entire $CD8^+$ (that is, $CD8^+$ T at L1) T cells due to the sparsity of TCR V genes (Supplementary Fig. 8). Please see the Supplementary Methods for details on the significance threshold.

Additionally, we performed genome-wide association analysis for integrative features representing repertoire variation for TCRs and BCRs. Integrative features included the top five PCs for V(D)J gene usage and mean and coefficient of variation for the length of CDR3 sequences and the frequency of SHM. SHM was calculated using the R package SHazaM (version 1.1.0)[66]. As potential biases in measuring SHM due to the complex structure of germline variants in VDJ genes could not be excluded, GWAS results of SHM should be interpreted carefully. For TRBD genes, we used *TRBD1* gene usage instead of PCs because they only have *TRBD1* and *TRBD2*. For BCR J genes, we used the top two PCs, as the number of these genes is fewer than six. The analysis was conducted based on a liner regression model for each feature normalized by rank-based inverse normal transformation with the same covariates using PLINK (version 2.00). For this analysis, we used $P = 2.0 \times 10^{-9}$ and 1.3×10^{-9} for the significance thresholds of the TCR and BCR repertoires, respectively, on the basis of Bonferroni correction on the number of features for a genome-wide significance threshold of $P = 5.0 \times 10^{-8}$. Due to sparsity, this analysis was applied only to $CD4^+$ T, $CD4_N$ and $CD4_{TCM}$ and $CD8^+$ T cells for TCRs and entire, type 1 naive, type 2 naive and memory B cells for BCRs.

Dynamic eQTL analysis with the pseudobulk approach

To identify dynamic genetic regulation of gene expression along continuous cell states, we investigated dynamic eQTLs in the myeloid cluster (that is, monocytes and dendritic cells) across the two gene modules 'HALLMARK_INFLAMMATORY_RESPONSE' and 'GOBP_RESPONSE_TO_INTERFERON_GAMMA'. We downloaded the two gene sets from MSigDB. Module scores were evaluated at the single-cell level using the AddModuleScore() function implemented in Seurat with default parameters. We divided the gene modules into ten windows containing roughly equal numbers of cells and averaged the expression of each gene per sample within each window separately for the two gene modules (Fig. 4a). Pseudobulk matrices per sample within each window with more than ten cells were retained in the analysis. Next, the averaged expression was normalized using inverse normal transformation, and gene expression PCs were calculated. To account for the higher correlation in expression values derived from the same individual at multiple gene module windows, we applied (1) linear and (2) quadratic mixed models, with individuals modeled as random intercepts as previously described[6,8]. We used these models to test for a significant interaction between genotypes and module categories as follows:

$$\begin{aligned}\mathrm{Exp} = {} & \beta_0 + \beta_g \times \mathrm{genotype} + \beta_a \times \mathrm{age} + \beta_{sex} \times \mathrm{sex} + \beta_v \times \mathrm{version} \\ & + \beta_{status} \times \mathrm{status} + \sum_{i=1}^{2} \beta_{gPC_i} \times \mathrm{gPC}_i + \sum_{j=1}^{15} \beta_{ePC_j} \times \mathrm{ePC}_j \\ & + \beta_m \times \mathrm{module} + \beta_{g\times m} \times \mathrm{genotype} \times \mathrm{module} + (1|\mathrm{sample})\end{aligned} \tag{1}$$

$$\begin{aligned}\mathrm{Exp} = {} & \beta_0 + \beta_g \times \mathrm{genotype} + \beta_a \times \mathrm{age} + \beta_{sex} \times \mathrm{sex} \\ & + \beta_v \times \mathrm{version} + \beta_{status} \times \mathrm{status} \\ & + \sum_{i=1}^{2} \beta_{gPC_i} \times \mathrm{gPC}_i + \sum_{j=1}^{15} \beta_{ePC_j} \times \mathrm{ePC}_j + \beta_m \times \mathrm{module} \\ & + \beta_{m^2} \times \mathrm{module}^2 \\ & + \beta_{g\times m} \times \mathrm{genotype} \times \mathrm{module} \\ & + \beta_{g\times m^2} \times \mathrm{genotype} \\ & \times \mathrm{module}^2 + (1|\mathrm{sample}).\end{aligned} \tag{2}$$

gPC represents the genotype PC and ePC represents the gene expression PC. In both cases, the null model was computed using the same parameters while excluding the interaction terms of the genotype × module category and the genotype × module category^2. *P* values were calculated by comparing each model to its respective null model using a likelihood ratio test. All models were implemented in R using the lmer() function. To minimize multiple-testing burden and focus on more robust candidates, we only applied this approach to SNP–gene pairs identified as significant lead eQTL variants by tensorQTL in at least one module window separately for the two gene modules. Next, we determined the top SNP per gene for each model and module. Finally, to correct for the number of genes tested and identify significant dynamic eGenes (deGenes), the FDR was calculated via the Benjamini–Hochberg method for the top associated SNP–gene pairs, setting an FDR threshold of 0.05. We also performed the same analysis across 'the type 1 ISG signature score (ISG score)' (ref. 7).

To find enriched pathways of module-shared and module-specific deGenes, we used the compareCluster function (fun = 'enrichGO', pvalueCutoff = 0.05, pAdjustMethod = 'BH', OrgDb = 'org.Hs.eg.db', ont = 'BP') of clusterProfiler (version 3.14.3)[67].

Epigenetic marker enrichment analysis for dynamic eQTLs

We evaluated the enrichment of significant dynamic eQTLs to epigenome marks[35] as we did for *cis*-eQTLs. We compared the enrichment of significant dynamic eQTLs and *cis*-eQTLs of the myeloid cluster (L1 and L2) in promoters and enhancers of monocytes in Roadmap separately for each model and module. For dynamic eQTLs, we only considered the top associated SNPs per gene analyzed in dynamic eQTLs for each model and module, and, for *cis*-eQTLs, we considered the top associated SNPs per gene in each cell type.

We selected the genes with significant dynamic eQTLs located in promoters or enhancers of monocytes in Roadmap for each module and also selected the genes with significant *cis*-eQTLs located in promoters and enhancers of monocytes among any cell type of the myeloid cluster (L1 and L2). Next, we identified genes with eQTLs located in functional regions specifically for dynamic eQTLs. To evaluate the enriched pathways of these genes for each module and functional region, we used the compareCluster function (fun = 'enrichGO', pvalueCutoff = 0.2, pAdjustMethod = 'BH', OrgDb = 'org.Hs.eg.db', ont = 'BP') of clusterProfiler (version 3.14.3)[67].

Dynamic eQTL analysis at single-cell resolution

We performed eQTL mapping at single-cell resolution using the NBME model[68]. To represent the continuous cell state of each cell, we used hPCs calculated with Harmony (version 0.1)[69] from the top 30 original PCs in the myeloid cluster. The negative binomial model was fitted with the following formula using the glmer.nb() function in the lme4 R library (version 1.1.32):

$$\begin{aligned}\mathrm{Exp}\,(\mathrm{UMI}) = {} & \beta_0 + \beta_\mathrm{g} \times \mathrm{genotype} + \beta_\mathrm{a} \\ & \times \mathrm{age} + \beta_\mathrm{sex} \times \mathrm{sex} + \beta_\mathrm{v} \times \mathrm{version} + \beta_\mathrm{status} \times \mathrm{status} \\ & + \beta_\mathrm{UMI} \times \log \mathrm{UMI} + \beta_\mathrm{MT} \times \mathrm{MT} \\ & + \sum_{i=1}^{2} \beta_{\mathrm{gPC}_i} \times \mathrm{gPC}_i + \sum_{j=1}^{15} \beta_{\mathrm{ePC}_j} \times \mathrm{ePC}_j + \sum_{k=1}^{15} \beta_{\mathrm{hPC}_k} \times \mathrm{hPC}_k \\ & + \sum_{k=1}^{15} \beta_{\mathrm{g}\times\mathrm{hPC}_k} \times \mathrm{genotype} \times \mathrm{hPC}_k + (1|\mathrm{sample})\,. \end{aligned} \quad (3)$$

MT represents percent mitochondrial genes, and ePC represents the gene expression PC of the raw data. To assess whether the eQTL is cell state dependent, we compared the full model (equation (3)) to the null model without interaction terms using the likelihood ratio test. To calculate eQTL strength across cell states, we combined the genotype main effect (β_g) with the interaction effects of each hPC, weighted by each cell's position along each hPC (equation (4)).

$$\beta_\mathrm{total} = \beta_\mathrm{g} + \sum_{k=1}^{15} \beta_{\mathrm{g}\times hPC_k} \times \mathrm{hPC}_k. \quad (4)$$

To assess which hPCs had a strong interaction effect with genotype, we calculated *P* values of interaction terms in the full model. To understand which cell state each PC represents, we investigated a loading of each gene along each PC in the SCT assay using the Loadings() function implemented in Seurat and determined the genes that showed strong correlation (>0.1 absolute value) for each PC. We performed pathway enrichment analysis of PC-associated genes using the compareCluster function (fun = 'enrichGO,' pvalueCutoff = 0.05, pAdjustMethod = 'BH', OrgDb = 'org.Hs.eg.db', ont = 'BP') of clusterProfiler (version 3.14.3)[67].

Colocalization analysis

We download GWAS summary statistics in an East Asian population of 13 complex immune traits[70–74] except that of hospitalized COVID-19, which was obtained through collaboration with the Japan COVID-19 Task Force (JCTF)[47,48] (Supplementary Table 18). GWAS summary statistics in hg37 were converted to hg38 ones using LiftOver. For evaluation of eQTL signal colocalization with GWAS signals, we applied coloc (version 5.2.3)[75]. We tested for 500-kb windows centered on significant GWAS lead variants (P value < 5×10^{-8}) outside the MHC region and considered PP.H4 > 0.8 as significant colocalization. If the GWAS lead variants were listed in the paper, we used them, and, if not, we manually defined them by clumping using PLINK version 1.9 software. In a heatmap (Extended Data Fig. 6b), the gene with the highest number of PP.H4 > 0.8 or, in case of a tie, the gene with the higher PP.H4 was presented.

Plasma protein expression measurements

We measured the expression of 2,925 plasma proteins using the Olink Explore 3072 platform for 227 samples (83 patients with COVID-19 and 144 healthy individuals) through collaboration with the JCTF[47,48]. The Olink Explore 3072 platform quantifies the expression of each protein in a normalized scale (normalized protein expression). Because the protein measurements were separated into three batches for logistical reasons, we bridge-normalized the normalized protein expression values using the OlinkAnalyze R package (version 3.4.1), using 16 intersecting samples as bridging samples. We excluded proteins with measurements below the detection limit in more than 10% of samples. As a result, 2,852 proteins were included in the analysis.

Polygenic risk score

We constructed multipopulation PRSs of hospitalized COVID-19 using PRS-CSx[45], combining the summary statistics of hospitalized COVID-19 (B2, n_case = 32,519) in Europeans from COVID-19 HGI (round 7)[46] and those of hospitalized COVID-19 (n_case = 2,948) in Japanese individuals from the JCTF[47,48] as the training data. We excluded the MHC region and set the φ value to 10^{-4}. Next, we calculated PRSs for OASIS samples and evaluated PRS performance with Nagelkerke's R^2.

To construct multipopulation PRSs of systemic lupus erythematosus[71,76], monocyte count[73,77] and lymphocyte count[73,77], we performed the same procedure as described for COVID-19. For rheumatoid arthritis[70], inflammatory bowel diseases[72], Crohn's disease[72] and ulcerative colitis[72], we used the publicly available multipopulation PRSs. We could not evaluate PRS performance in the OASIS cohort except for the COVID-19 PRS.

Differential gene and protein expression analysis with polygenic risk score

We divided each PRS into four quantiles in the OASIS cohort and treated them as ordinal variables 1 (lowest risk) to 4 (highest risk). DEG analysis among four quantiles of PRS was performed using the pseudobulk approach separately for patients with COVID-19 and healthy individuals. Pseudobulk matrices were created by aggregating gene counts for each cell type (L1) in each sample. Genes were considered for the analysis if they were expressed in more than 10% of cells per cell type. This analysis was performed using edgeR (version 3.32.0)[78]. We included age, sex, 10x chemistry and severity (severity was included only in patients with COVID-19) in the model as covariates. DEP analysis among four quantiles of PRSs was performed using a linear regression model with age, sex, batches and severity (severity was included only in patients with COVID-19) as covariates. Statistically significant DEGs and DEPs were defined with FDR < 0.1.

We permuted four quantiles of PRSs across samples 1,000 times in each cell type of scRNA-seq and proteomics data and assessed how often the observed number of DEGs and DEPs would be detected.

To investigate whether DEGs and DEPs between patients with COVID-19 and healthy individuals reflected those of PRSs specific to patients with COVID-19, DEG and DEP analysis between cases and controls were performed with the same method described above except that statistically significant DEGs were defined with FDR < 0.05 and absolute fold change > 2.

Detection of somatic events using genomics data

For the detection of mCAs and LOY from SNP array data, we used the MoChA pipeline (18 May 2022)[14,49]. In brief, IDAT genotype intensity data were converted to GTC genotype files and then transformed into VCF files with $\log_2$ (R ratio) and B allele frequency values to estimate total and relative allelic intensities, respectively. Using the phase information and $\log_2$ (R ratio) and B allele frequency values, MoChA was used to detect mCAs and LOY. From the resulting candidates, calls flagged as germline copy number polymorphisms and calls that were likely germline duplications were removed. For the LOY candidates, XXY and XXX samples were filtered out. We additionally removed unclassifiable calls and calls with lower cell fraction (<0.01).

We used a previously reported GATK-based pipeline[52] to detect mt-heteroplasmy from WGS data. To call homoplasmic and heteroplasmic variants in mtDNA (the circular genome including 16,569 bp), we used a mitochondrial mode implemented in GATK Mutect2. Variants with VAF > 0.9 were defined as homoplasmic, and we analyzed mt-heteroplasmy satisfying VAF > 0.1.

Single-cell deconvolution of somatic events

We applied Numbat (version 1.3.0)[50] to scRNA-seq data to detect clones with mCAs found in SNP array data. Numbat integrated haplotype information with allele and expression signals from scRNA-seq data to enhance detection of mCAs including CN-LOHs. Expression references of the COVID-19 and healthy samples were generated from our scRNA-seq data. We specified the profiles of the genotype-based mCAs and ran Numbat with parameters init_k = 10 and max_entropy = 1.0.

Using scRNA-seq data from all male samples, we defined cells with LOY as cells without expression from the male-specific region of the Y chromosome[22]. As single-cell-based LOY, we then classified the male samples as LOY high (cells with LOY > 5%) or LOY low (cells with LOY < 5%). The association between LOY and COVID-19 risk was evaluated using univariate logistic regression and multivariate logistic regression adjusted for age.

For mt-heteroplasmy detected in WGS data, we piled up raw reads from the scRNA-seq BAM files and generated SNP-by-cell matrices using cellSNP-lite (version 1.2.2)[79]. Next, we performed clonal assignment of each cell using the BinomMixtureVB function (binomial mixture model) implemented in vireoSNP (version 0.5.8)[80]. We included samples in which more than 80% of cells showed posterior clonal assignment probability > 0.8.

DEG analysis between somatically mutated and normal cells

DEG analysis for CNAs was conducted for mutated and normal cells from the same samples in CH01 and CH05. We used MAST (version 1.20.0)[81] implemented in Seurat, and DEGs were considered significant if they satisfied FDR (adjusted *P* values via the Benjamini–Hochberg method) < 0.05 and $\log_2$ (fold change) > 0.25 (in CH01) or 0.5 (in CH05). For significant DEGs, pathway enrichment analysis was performed using the enrichGO function of clusterProfiler[67] with parameters OrgDb = 'org.Hs.eg.db', ont = 'BP', pvalueCutoff = 0.05 and pAdjustMethod = 'BH'. In the CH05 analysis, immunoglobulin genes were excluded from the pathway analysis.

DEG analysis between cells with LOY and normal cells was performed using a pseudobulk approach. Pseudobulk matrices were created by aggregating gene counts for each cell type within cells with LOY or normal cells in each sample. Genes were included if they were expressed in more than 10% of cells. We included samples with more than nine cells with LOY and more than nine normal cells in the cell type. This analysis was performed using a linear mixed model implemented in the lme4 R library (version 1.1.32) with fixed effects for age and 10x chemistry and random effects for sample. DEGs were considered significant if they satisfied FDR (adjusted *P* values via the Benjamini–Hochberg method) < 0.01 and $\log_2$ (fold change) > 0.25. Pathway analysis was performed using the same procedure as for mCAs.

Differential abundance analysis of LOY

We used Milo (version 1.2.0)[37] to test for the differential abundance of cells within defined neighborhoods between LOY-high and LOY-low samples. This analysis was performed in the same manner as the metagenomic analysis. Comparisons of cell proportions between LOY-high and LOY-low samples were carried out using two-sided Wilcoxon rank-sum test.

Statistics and reproducibility

No statistical methods were used to predetermine sample sizes. Data distribution was assumed to be normal, but this was not formally tested. No data were excluded from the analyses. We did not use any study design that required randomization or blinding.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw sequencing data of scRNA-seq and the protein expression matrix are available at the Japanese Genotype–phenotype Archive (JGA) with accession codes JGAS000783 and JGAD000925. A part of the raw scRNA-seq data ($n_{\text{COVID-19}} = 73$, $n_{\text{control}} = 75$)[44,47] has already been deposited and is available under controlled access at JGA with accession codes JGAS000593, JGAS000543, JGAD000662 and JGAD000722. All raw scRNA-seq data can also be accessed by applying at the NBDC with the accession code hum0197. Participant genotype data are available at the European Genome–Phenome Archive with the accession code EGAS00001008016. sc-eQTL summary statistics are available at the NBDC under accession ID hum0197 and are also available in an interactive browser at https://japan-omics.jp/. OneK1K genotype data were obtained from the Gene Expression Omnibus (GSE196830). Processed OneK1K scRNA-seq data (h5ad file) were obtained from the Human Cell Atlas (https://explore.data.humancellatlas.org/projects/f2078d5f-2e7d-4844-8552-f7c41a231e52). OneK1K eQTL summary statistics were obtained from https://onek1k.org/.

Code availability

The codes used in this study are shared on GitHub (https://github.com/REdahiro/OASIS_project and https://github.com/tatsuhikonaito/OASIS_HLATCR) and have been deposited at Zenodo (https://doi.org/10.5281/zenodo.15877644 (ref. 82) and https://doi.org/10.5281/zenodo.14991132 (ref. 83)).

References

54. World Health Organization. *COVID-19 Therapeutic Trial Synopsis* (World Health Organization, 2020).
55. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
56. Lun, A. T. L., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17**, 75 (2016).
57. Taylor-Weiner, A. et al. Scaling computational genomics to millions of individuals with GPUs. *Genome Biol.* **20**, 228 (2019).
58. Tomofuji, Y. et al. Metagenome-wide association study revealed disease-specific landscape of the gut microbiome of systemic lupus erythematosus in Japanese. *Ann. Rheum. Dis.* **80**, 1575–1583 (2021).
59. Tomofuji, Y. et al. Prokaryotic and viral genomes recovered from 787 Japanese gut metagenomes revealed microbial features linked to diets, populations, and diseases. *Cell Genom.* **2**, 100219 (2022).
60. Sturm, G. et al. Scirpy: a Scanpy extension for analyzing single-cell T-cell receptor-sequencing data. *Bioinformatics* **36**, 4817–4818 (2020).
61. Hirata, J. et al. Genetic and phenotypic landscape of the major histocompatibilty complex region in the Japanese population. *Nat. Genet.* **51**, 470–480 (2019).
62. Naito, T. et al. A deep learning method for HLA imputation and trans-ethnic MHC fine-mapping of type 1 diabetes. *Nat. Commun.* **12**, 1639 (2021).
63. Naito, T. & Okada, Y. HLA imputation and its application to genetic and molecular fine-mapping of the MHC region in autoimmune diseases. *Semin. Immunopathol.* **44**, 15–28 (2022).
64. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
65. Kanai, M., Tanaka, T. & Okada, Y. Empirical estimation of genome-wide significance thresholds based on the 1000 Genomes Project data set. *J. Hum. Genet.* **61**, 861–866 (2016).
66. Gupta, N. T. et al. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics* **31**, 3356–3358 (2015).

67. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16**, 284–287 (2012).
68. Kang, J. B. et al. Mapping the dynamic genetic regulatory architecture of HLA genes at single-cell resolution. *Nat. Genet.* **55**, 2255–2268 (2023).
69. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
70. Ishigaki, K. et al. Multi-ancestry genome-wide association analyses identify novel genetic mechanisms in rheumatoid arthritis. *Nat. Genet.* **54**, 1640–1651 (2022).
71. Wang, Y. F. et al. Identification of 38 novel loci for systemic lupus erythematosus and genetic heterogeneity between ancestral groups. *Nat. Commun.* **12**, 772 (2021).
72. Liu, Z. et al. Genetic architecture of the inflammatory bowel diseases across East Asian and European ancestries. *Nat. Genet.* **55**, 796–806 (2023).
73. Sakaue, S. et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53**, 1415–1424 (2021).
74. Shirai, Y. et al. Multi-trait and cross-population genome-wide association studies across autoimmune and allergic diseases identify shared and distinct genetic component. *Ann. Rheum. Dis.* **81**, 1301–1312 (2022).
75. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
76. Bentham, J. et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat. Genet.* **47**, 1457–1464 (2015).
77. Vuckovic, D. et al. The polygenic and monogenic basis of blood traits and diseases. *Cell* **182**, 1214–1231 (2020).
78. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
79. Huang, X. & Huang, Y. Cellsnp-lite: an efficient tool for genotyping single cells. *Bioinformatics* **37**, 4569–4571 (2021).
80. Huang, Y., McCarthy, D. J. & Stegle, O. Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. *Genome Biol.* **20**, 273 (2019).
81. Finak, G. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).
82. Edahiro, R. REdahiro/OASIS_project: v1.1. *Zenodo* https://doi.org/10.5281/zenodo.15877644 (2025).
83. Naito, T. tatsuhikonaito/OASIS_HLATCR: v1.0.0. *Zenodo* https://doi.org/10.5281/zenodo.14991132 (2025).

Acknowledgements

We sincerely thank all participants involved in this study. We thank and acknowledge T. Shiroyama, K. Miyake, Y. Suga and Y. Naito for contributing to clinical practice. We are grateful to Y. Yamaguchi, T. Murakami, Y. Noda, T. Niitsu, Y. Adachi, T. Enomoto, S. Amiya, R. Hara, M. Yamamoto, T. Kuge, K. Matsumoto, M. Yoneda, Y. Yamamoto, Y. Yoshimine, S. Minoda, T. Hirayama, K. Funakoshi, Y. Okita, S. Kawada, D. Nakatsubo, T. Tada, M. Okamoto and H. Shimagami for contributing to clinical practice and sample collection. We thank all members of the JCTF for their support and sharing GWAS summary statistics. This study was supported by AMED (JP23km0405217, JP23ek0109594, JP23ek0410113, JP23kk0305022, JP223fa627001, JP223fa627002, JP223fa627010, JP233fa627011, JP23zf0127008, JP23tm0524002, JP24wm0625504, JP24gm1810011 to Y. Okada and 24ek0210207h0001 to R.E.), JST Moonshot R&D (JPMJMS2021, JPMJMS2024 to Y. Okada), JSPS KAKENHI (25H01057, 24K02697 to Y. Okada), AMED–CREST (22gm1810003h0001 to A. Kumanogoh and JP24gm1810012 to R.E.), a Japan Foundation for Applied Enzymology grant, the Takeda Science Foundation, Ono Pharmaceutical Foundation for Oncology, Immunology, and Neurology, the Bioinformatics Initiative of Osaka University Graduate School of Medicine, the Institute for Open and Transdisciplinary Research Initiatives, CiDER and CAMaD, Osaka University and the RIKEN TRIP initiative (AGIS). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

R.E., G.S., A.K. and Y. Okada conceived and designed the study; R.E., Y. Shirai, Y. Tomofuji, T. Kishikawa, K. Suzuki, Y.-C.L., D. Motooka, A. Takuwa, D.O. and S. Sakakibara performed the experiments; R.E., G.S., Y. Shirai, R. Saiki, T. Naito, K.S., Y. Tomofuji, K. Yamamoto, S. Namba, N.S., G.N. (Japan COVID-19 Task Force), S.S. and Y. Okada conducted data analysis; R.E., G.S., T. Naito, K.S., Y. Tomofuji, S. Sakakibara, A.K. and Y. Okada wrote the manuscript; H. Hirata and Y. Takeda took care of patients and provided clinical information; Q.S.W., Y. Takahashi, T. Hasegawa, H. Tanaka, S. Azekawa, H. Namkoong, R. Koike, A. Kimura, S. Imoto, S. Miyano, T. Kanai, K. Fukunaga, M. Uemura, T. Morita, Y. Kato, H.E. and S. Ogawa provided intellectual input throughout the study, provided comments and helped edit the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-025-02266-3.

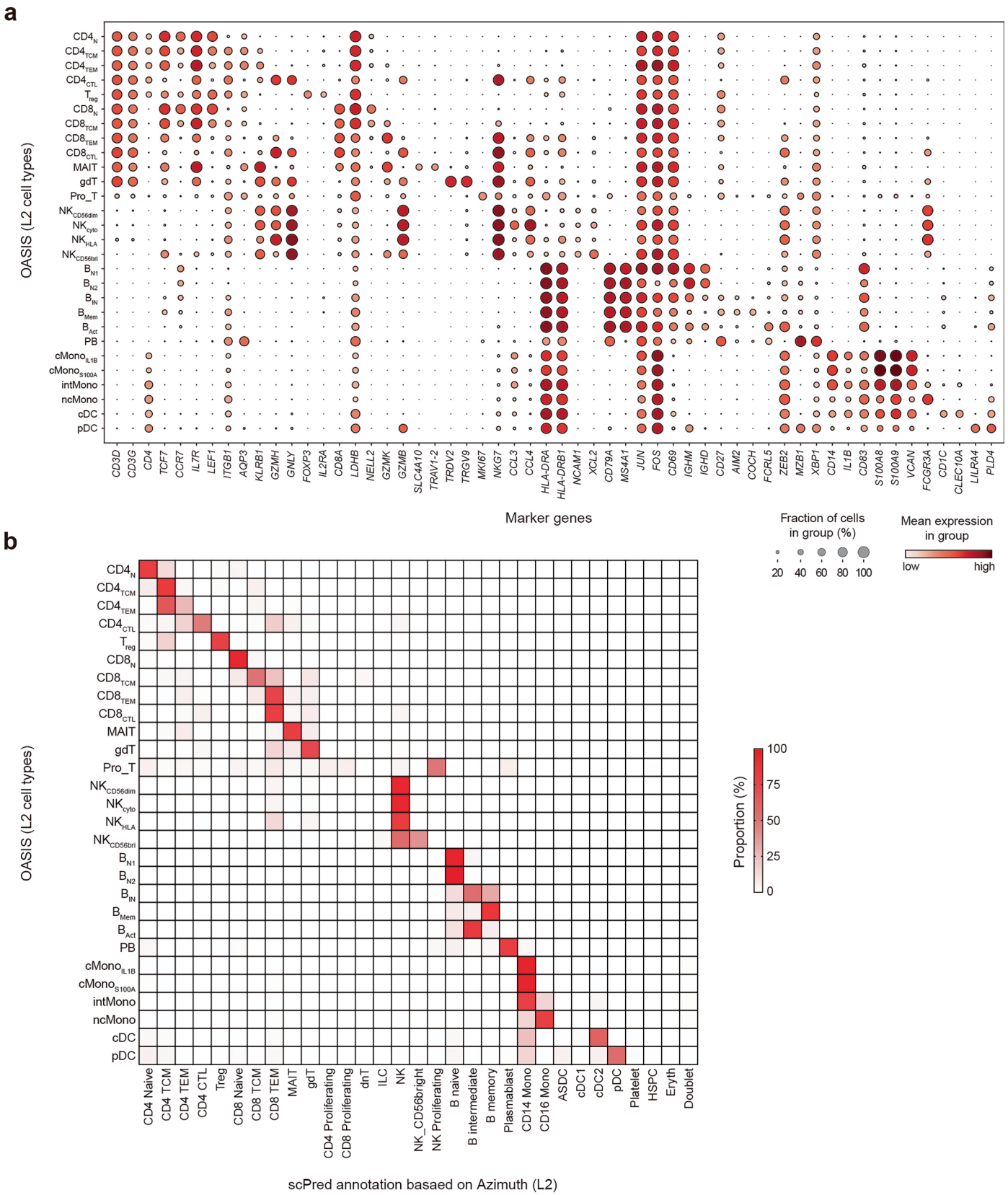
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-025-02266-3.

Correspondence and requests for materials should be addressed to Ryuya Edahiro, Atsushi Kumanogoh or Yukinori Okada.

Peer review information *Nature Genetics* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

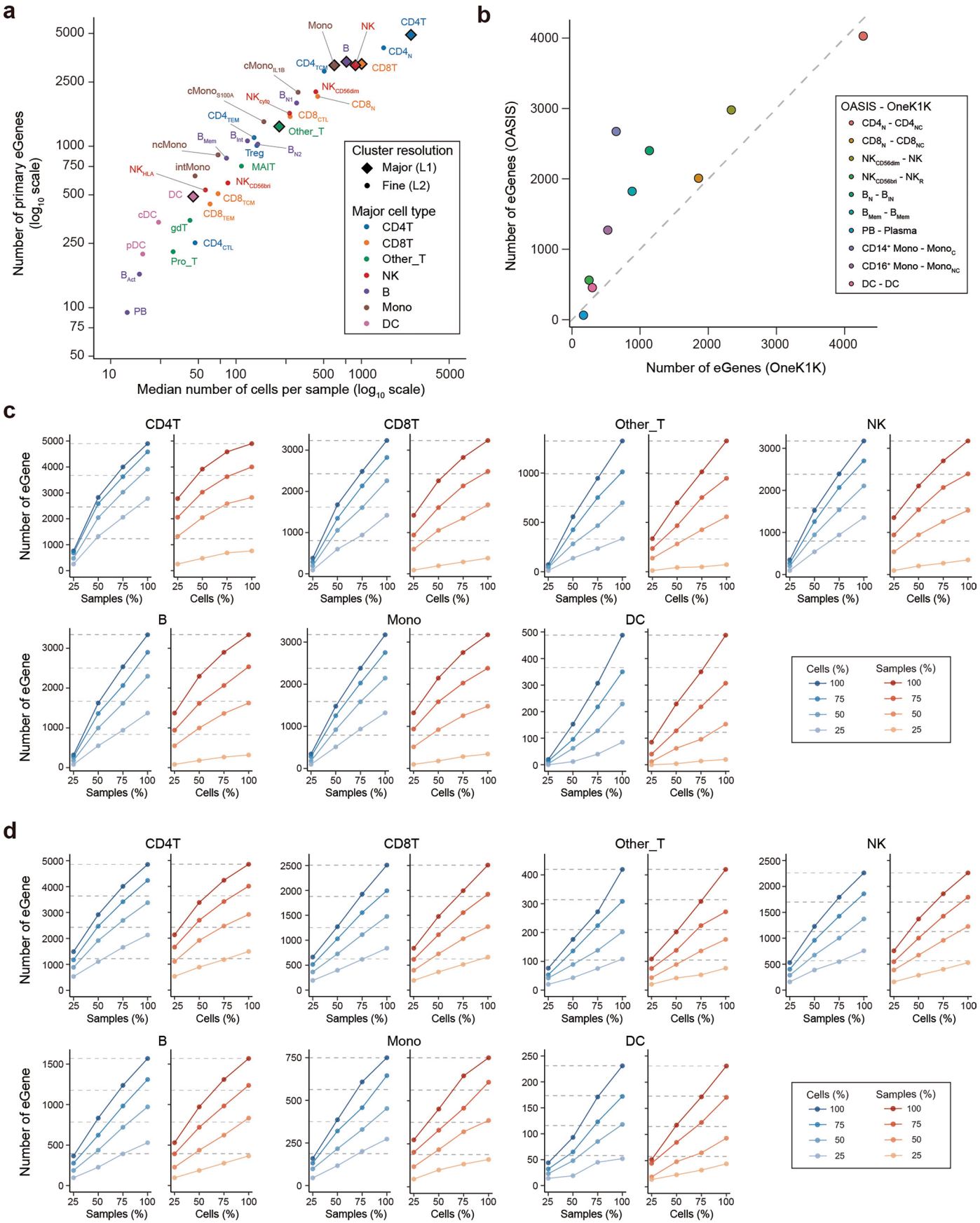
Reprints and permissions information is available at www.nature.com/reprints.

https://doi.org/10.1038/s41588-025-02266-3



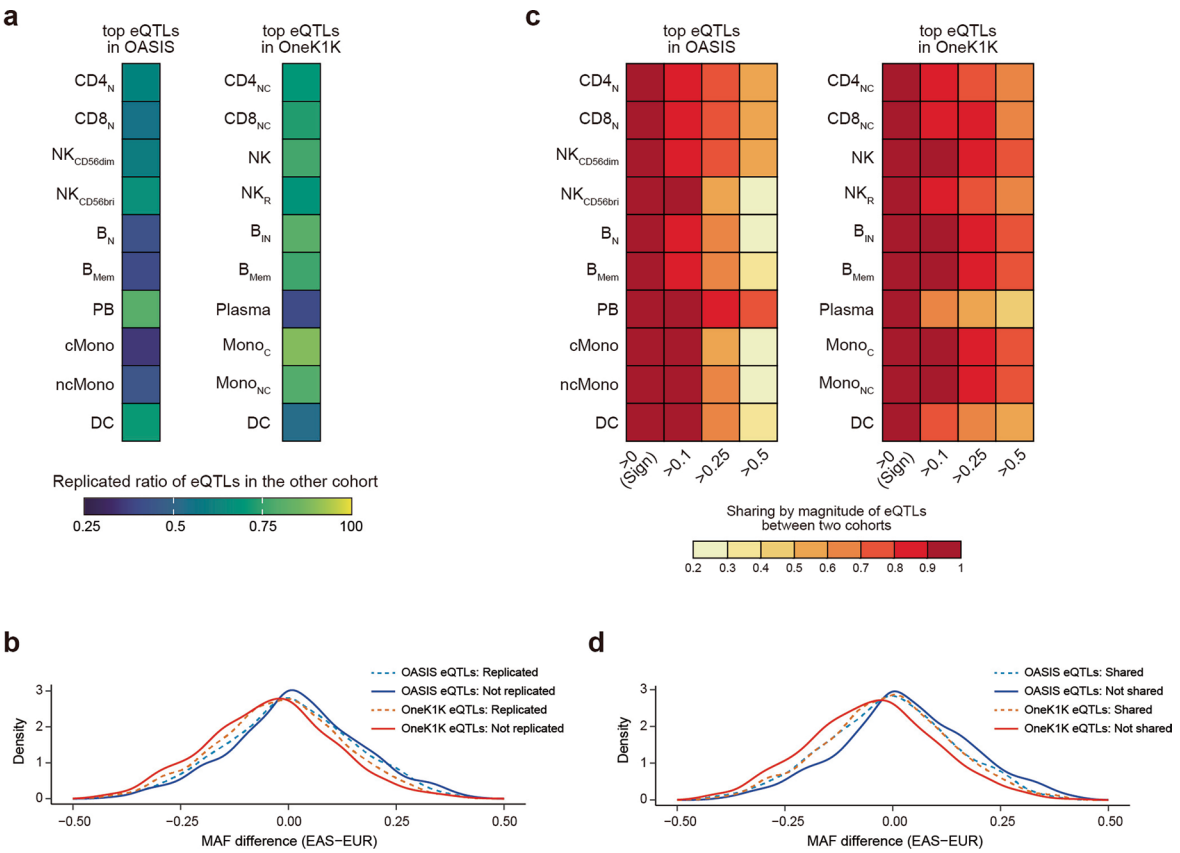
Extended Data Fig. 1 | Expression of cell type marker genes and reference-based annotation. (**a**) Dotplot showing the RNA expression of marker genes of L2 cell types. (**b**) Comparison of our manual PBMC annotation (L2 cell types) vs an automated annotation performed by scPred[32] with Azimuth[27] (L2) as a reference.

https://doi.org/10.1038/s41588-025-02266-3



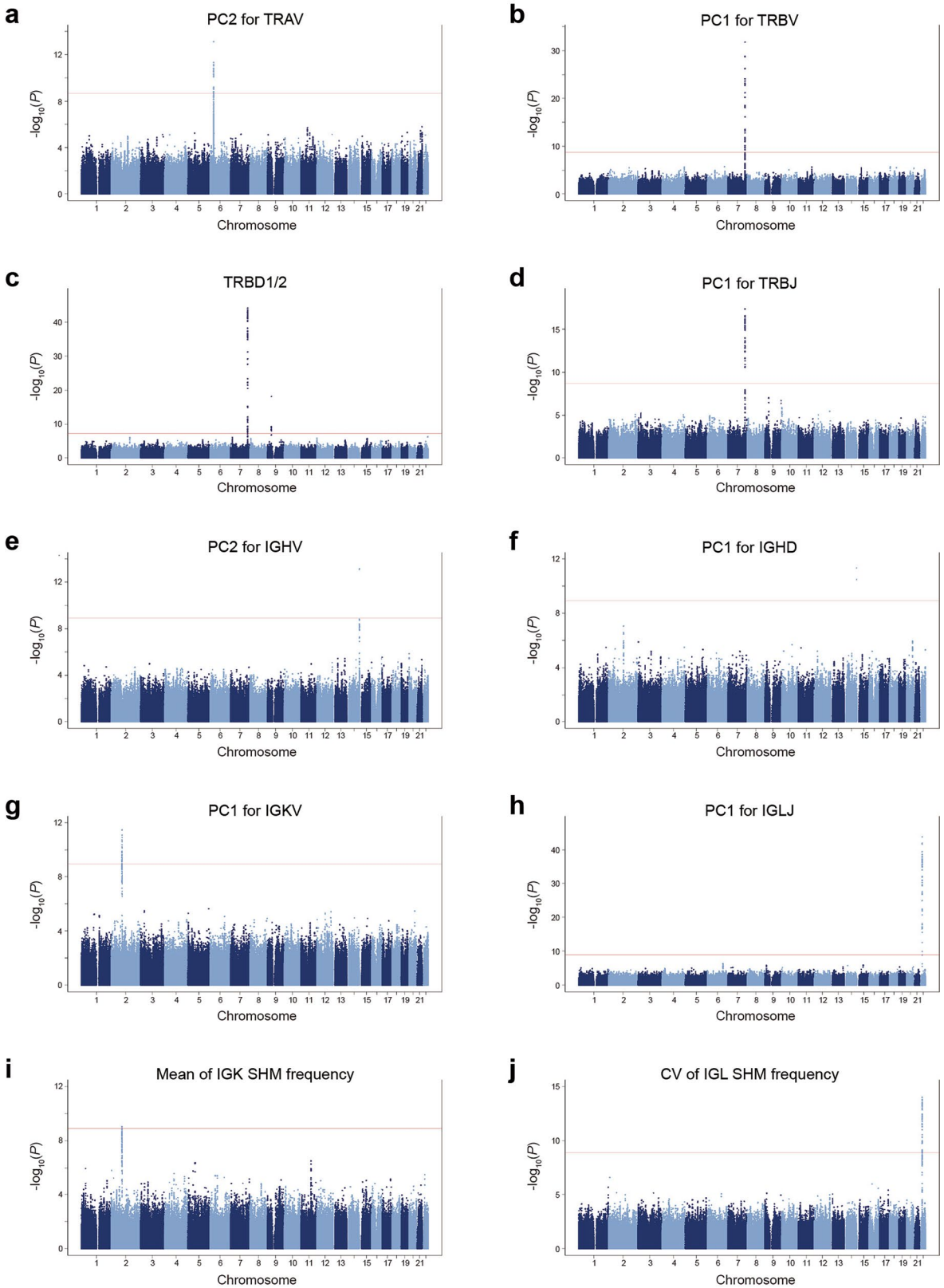
Extended Data Fig. 2 | The effect of sample sizes and cell counts per sample on eQTL discovery in OASIS and OneK1K. (**a**) Scatter plot depicting the correlation between median number of cells per sample and number of detected significant eGenes in each cell type of OASIS. Both x-axis and y-axis are on $\log_{10}$ scale. (**b**) Co-plots of the number of eGenes between OASIS and OneK1K[6] in corresponding cell types. (**c**,**d**) The number of eGenes by down-sampling for combinations of sample sizes and cell counts per sample in OASIS (**c**) and OneK1K (**d**) at L1 level.

https://doi.org/10.1038/s41588-025-02266-3



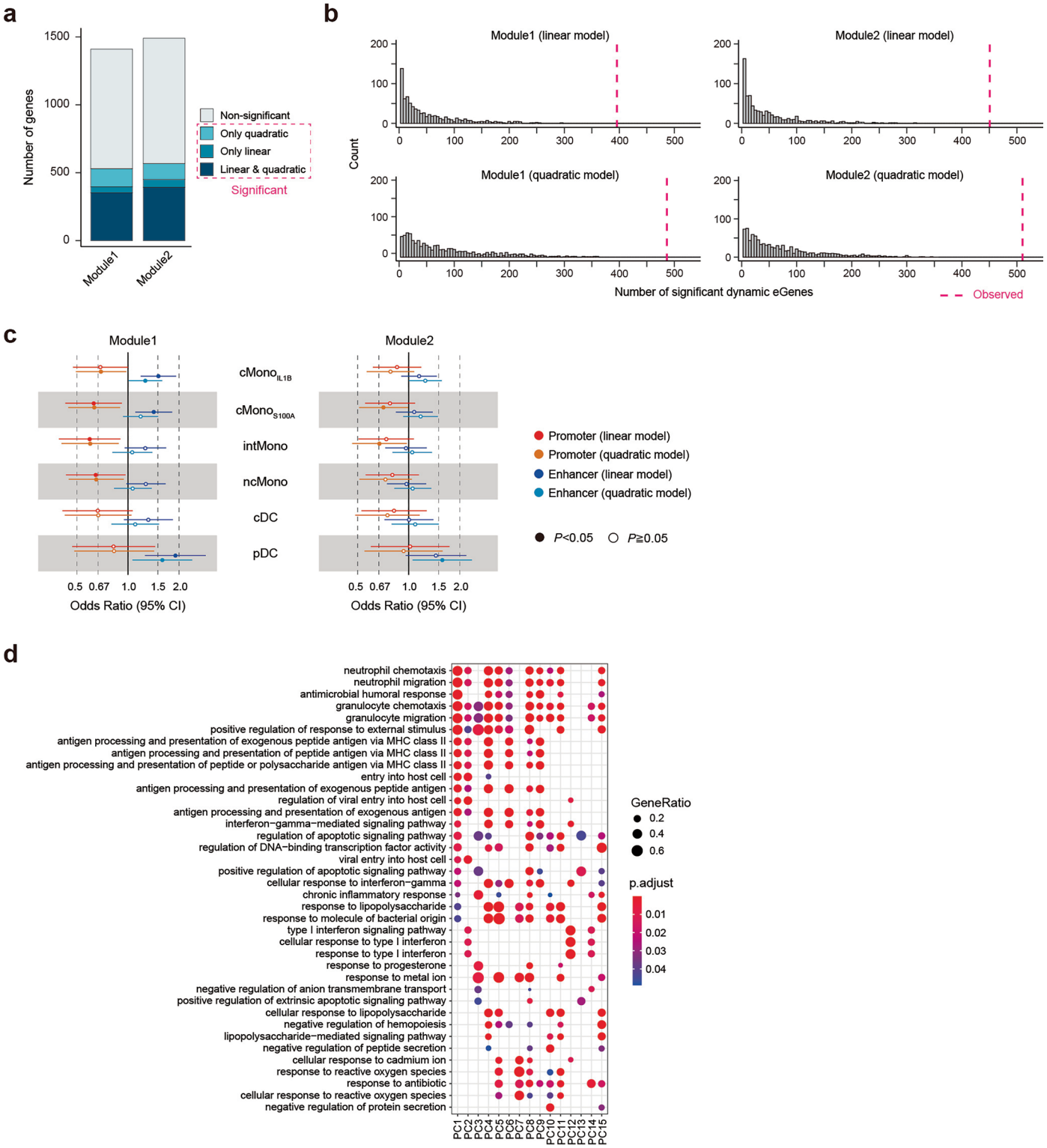
Extended Data Fig. 3 | Comparison of cis-eQTLs between OASIS and OneK1K in the corresponding 10 cell type pairs. (**a**) The replicated ratio of eQTLs between the two cohorts in the corresponding 10 cell type pairs. (**b**) Distribution of difference of minor allele frequency (MAF) between East Asian (EAS) and European (EUR) of cis-eQTLs according to with/without cross-population eQTL replication for each cohort in naive CD4$^+$ T cells. (**c**) Pairwise sharing by magnitude of eQTLs between the two cohorts in the corresponding 10 cell type pairs according to four thresholds of factor. (**d**) Distribution of difference of MAF between EAS and EUR of cis-eQTLs according to with/without cross-population eQTL sharing for each cohort in naive CD4$^+$ T cells. MAF from both ancestries was calculated from high-coverage 1000 Genome Project phase 3 in (**b**) and (**d**).

https://doi.org/10.1038/s41588-025-02266-3



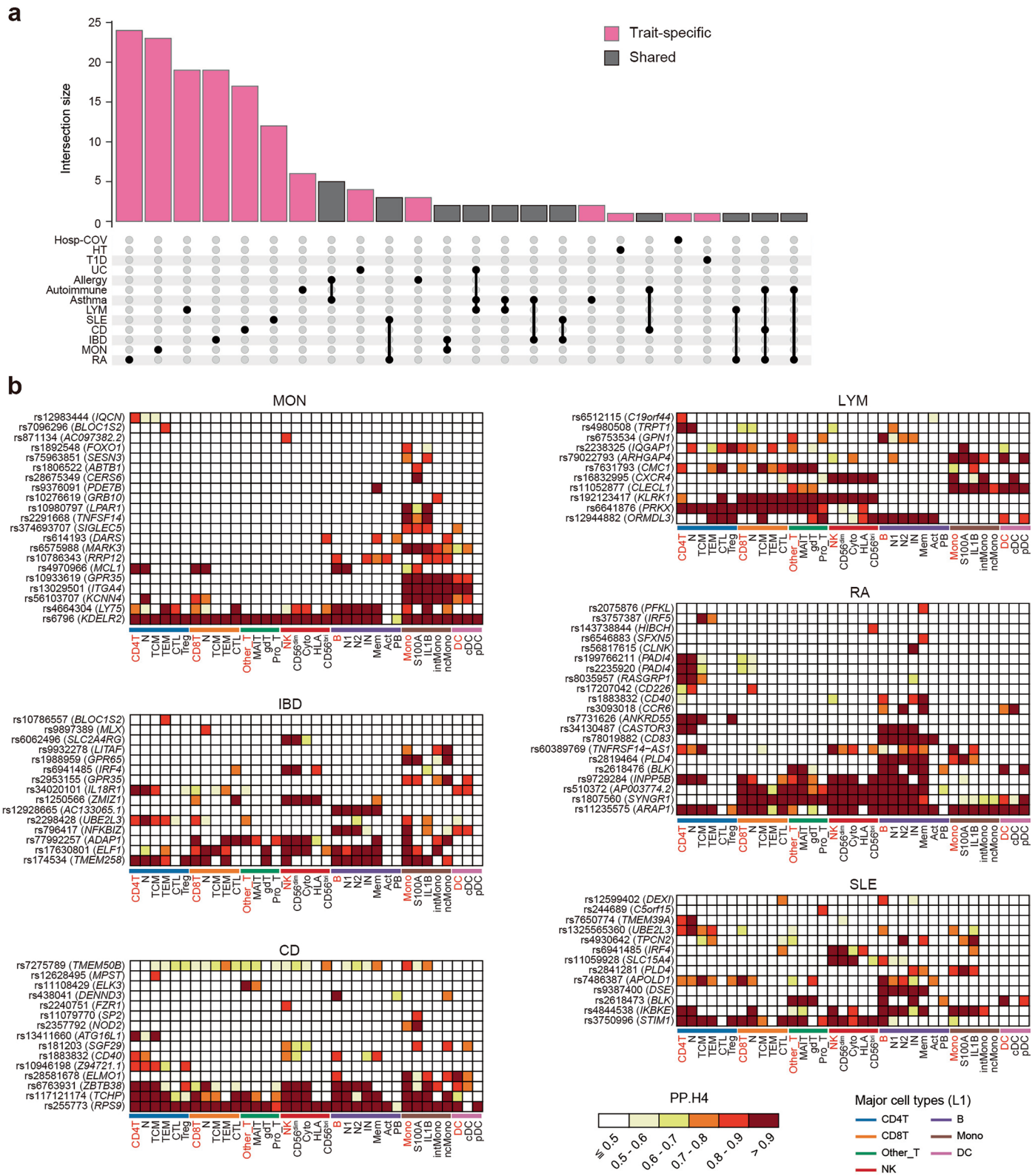
Extended Data Fig. 4 | Genome-wide association analysis for TCR/BCR repertoire features. Manhattan plots for the representative results for genome-wide association analysis for TCR/BCR repertoire features, including PC2 for TRAV gene usage (**a**), PC1 for TRBV gene usage (**b**), TRBD1/2 usage (**c**), and PC1 for TRBJ gene usage (**d**) as TCR features in $CD4^+$ T cells and PC2 for IGHV gene usage (**e**), PC1 for IGHD gene usage (**f**), PC1 for IGKV PC1 (**g**), and PC1 for IGLJ gene usage (**h**) in entire B cells and mean of somatic hypermutation (SHM) in IGK (**i**) and CV of SHM in IGL (**J**) in naive1 B cells as BCR features. The red horizontal line indicates the study-wide significance threshold ($P = 2.0 \times 10^{-9}$ and 1.6×10^{-9} for TCR and BCR repertoire features, respectively). Uncorrected *P* values from the GWAS analysis are shown in (**a**-**j**).

https://doi.org/10.1038/s41588-025-02266-3



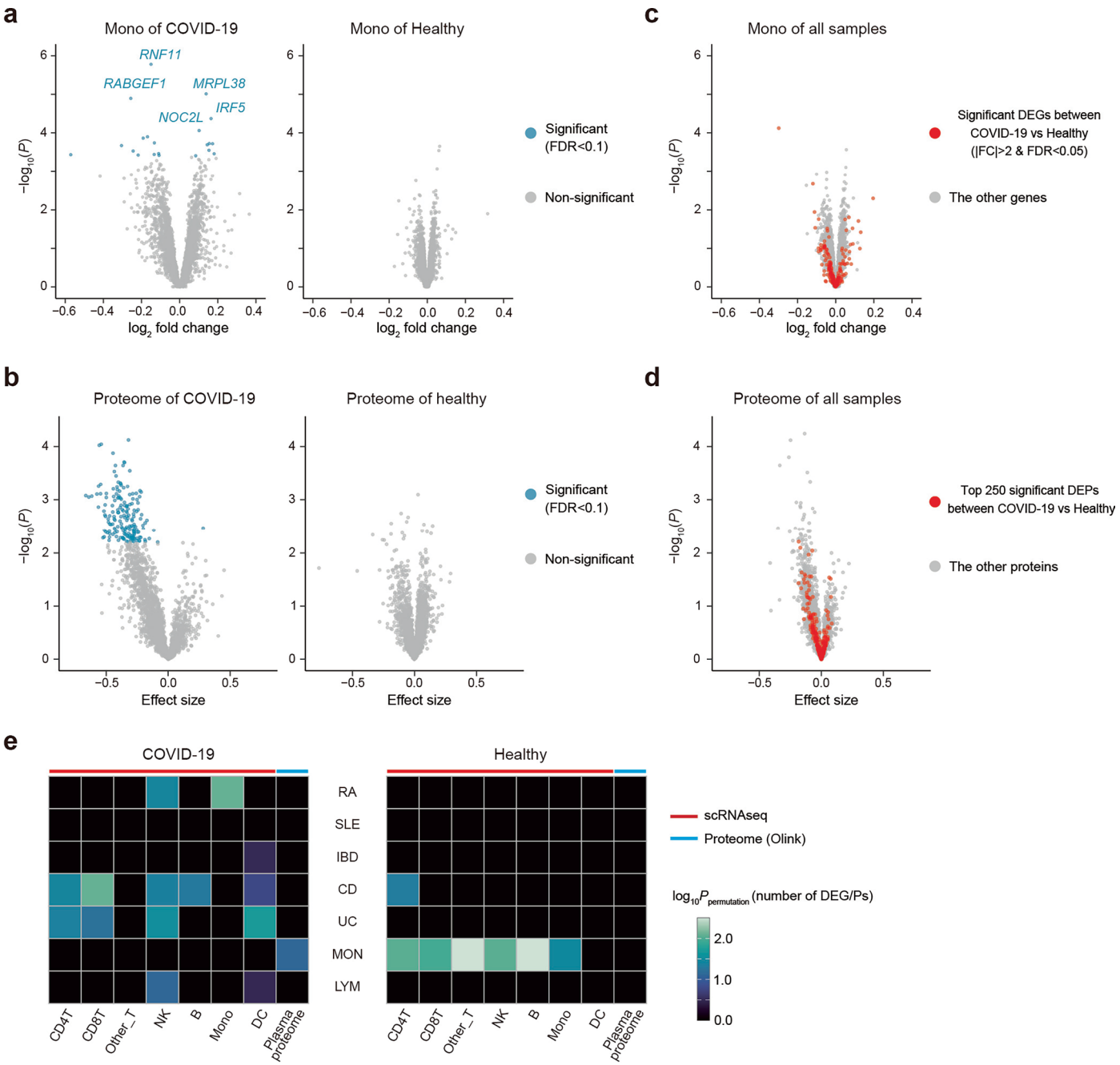
Extended Data Fig. 5 | Properties of dynamic eQTLs and pathways representing cell-state of each PC. (**a**) Number of genes with a significant or non-significant genotype-module interaction in a linear or quadratic mixed model for the two gene modules. (**b**) Distribution of the number of significant dynamic eGenes in 1,000 times permuted data for each combination of modules and analysis models. Pink dotted line represents the number of dynamic eGenes observed. (**c**) Forest plots showing odds ratios of overlap of dynamic eQTLs with functional regions in Monocytes from Roadmap project[35] compared to cis-eQTLs in L2 cell types of myeloid cluster for each combination of functional region and analysis model. Dots represent the odds ratios and bars represent the 95% confidence intervals. Two-sided *P* values are uncorrected. (**d**) The top 3 enriched pathways of the genes which were correlated with each PCs (> 0.1 absolute value). Dot color indicates the statistical significance of the enrichment (adjusted *P* values via the Benjamini–Hochberg method), and dot size represents gene ratio annotated to each term.

https://doi.org/10.1038/s41588-025-02266-3



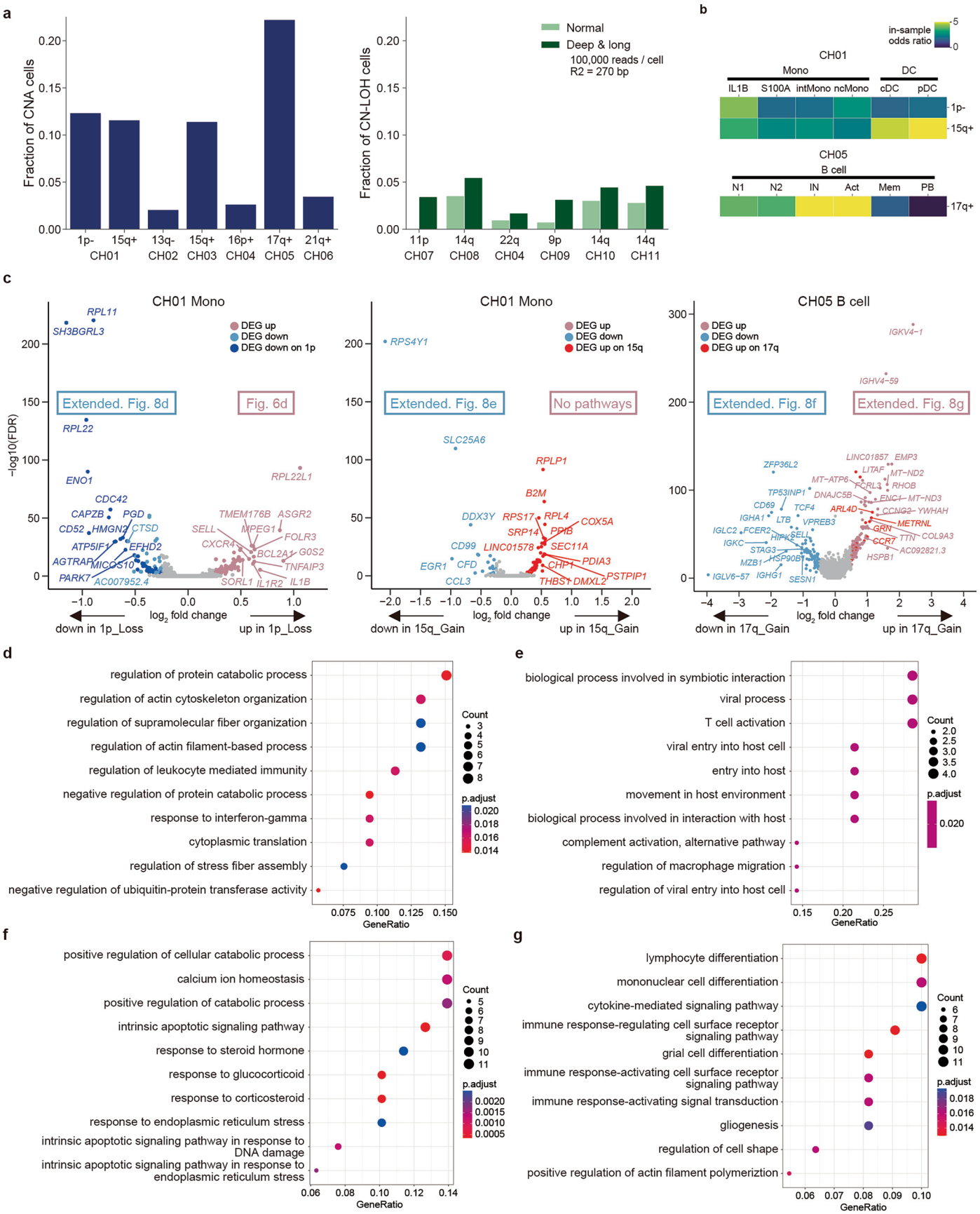
Extended Data Fig. 6 | Colocalizations of GWASs in East Asian population and cis-eQTLs in OASIS. (**a**) Number of colocalizing eGenes shared among different autoimmune and blood-related traits. Bar plots indicating the number of genes in each set. eGenes colocalizing in only one trait are colored by pink, and those in multiple traits are connected by lines and colored by gray. (**b**) Colocalizations of GWAS variants in East Asian population and cis-eQTLs in OASIS for 6 traits with more than 10 colocalizations. Heatmap depicting PP.H4 from coloc[75]. Cell types of major cell type (L1) are colored by red. GWAS abbreviations: CD, Crohn's disease; Hosp-COV, hospitalized COVID-19; HT, hyperthyroidism; IBD, inflammatory bowel diseases; LYM, lymphocyte count; MON, monocyte count; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; T1D, type 1 diabetes; UC, ulcerative colitis.

https://doi.org/10.1038/s41588-025-02266-3



Extended Data Fig. 7 | The effects of polygenic risk score on transcriptome and proteome. (**a**) The differential gene expression analysis with 4 quantiles of hospitalized COVID-19 polygenic risk score (PRS) in monocytes separately for COVID-19 patients and healthy subjects. Differentially expressed genes (DEGs; FDR < 0.1) are colored in light blue and labeled by gene symbols if *P* values < 1×10^{-4}. (**b**) The differential protein expression analysis with 4 quantiles of hospitalized COVID-19 PRS in plasma proteome separately for COVID-19 patients and healthy subjects. Differentially expressed proteins (DEPs; FDR < 0.1) are colored in light blue. (**c**) The differential gene expression analysis with 4 quantiles of hospitalized COVID-19 PRS in monocytes for all samples. DEGs between patients and healthy controls (FDR < 0.05 and absolute fold change (FC) > 2) are colored in red. (**d**) The differential protein expression analysis with 4 quantiles of hospitalized COVID-19 PRS in plasma proteome for all samples. Top 250 significant DEPs between patients and healthy controls (FDR < 4.0×10^{-22}) are colored in red. Uncorrected *P* values are shown in (**a**-**d**). (**e**) The effect of PRS for 7 autoimmune and blood-related traits on transcriptomics and proteomics levels separately in patients with COVID-19 and healthy subjects. One-sided *P* values were calculated with 1,000 permutations for number of DEGs and DEPs. GWAS abbreviations: CD, Crohn's disease; IBD, inflammatory bowel diseases; LYM, lymphocyte count; MON, monocyte count; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; UC, ulcerative colitis.

https://doi.org/10.1038/s41588-025-02266-3

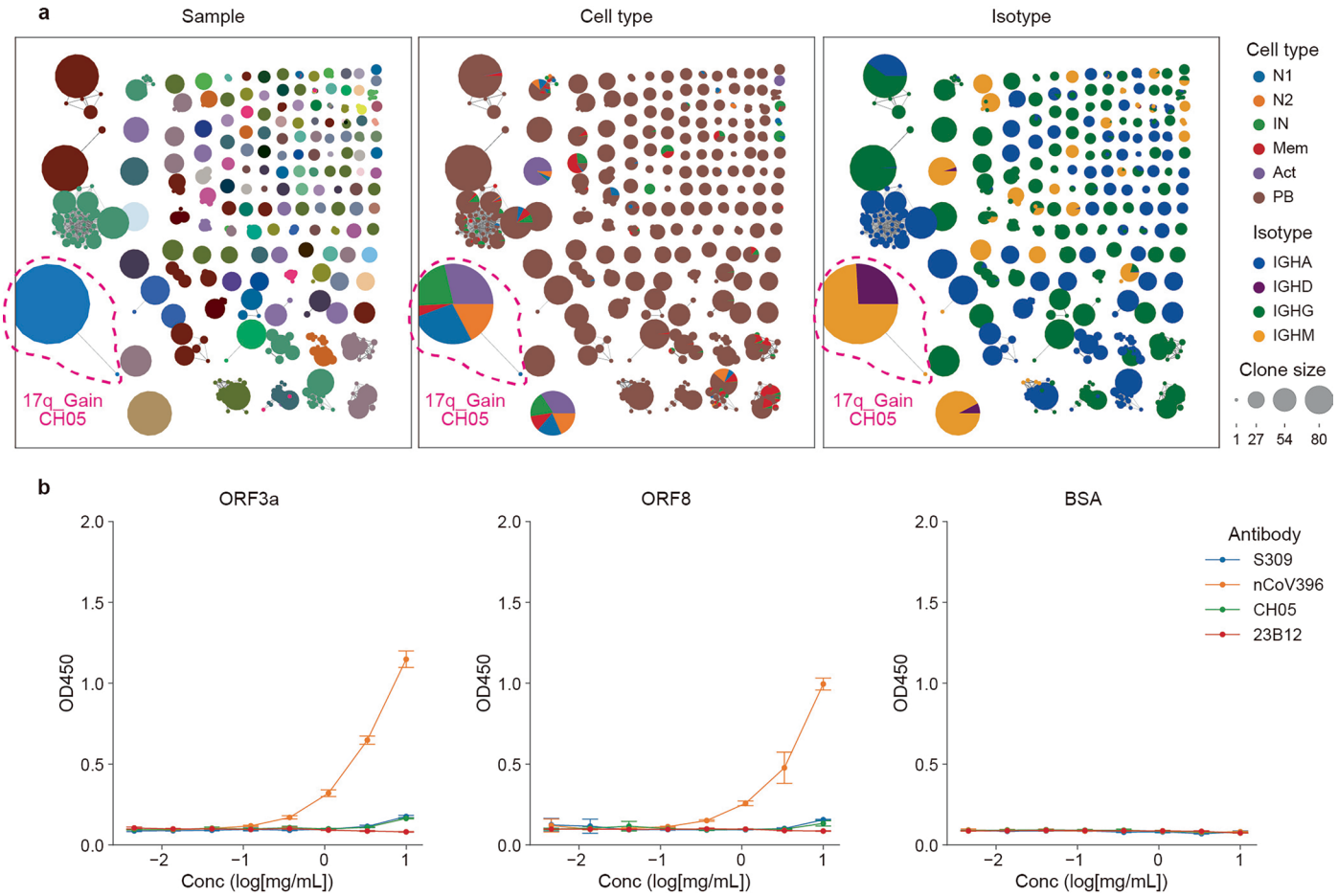


Extended Data Fig. 8 | See next page for caption.

https://doi.org/10.1038/s41588-025-02266-3

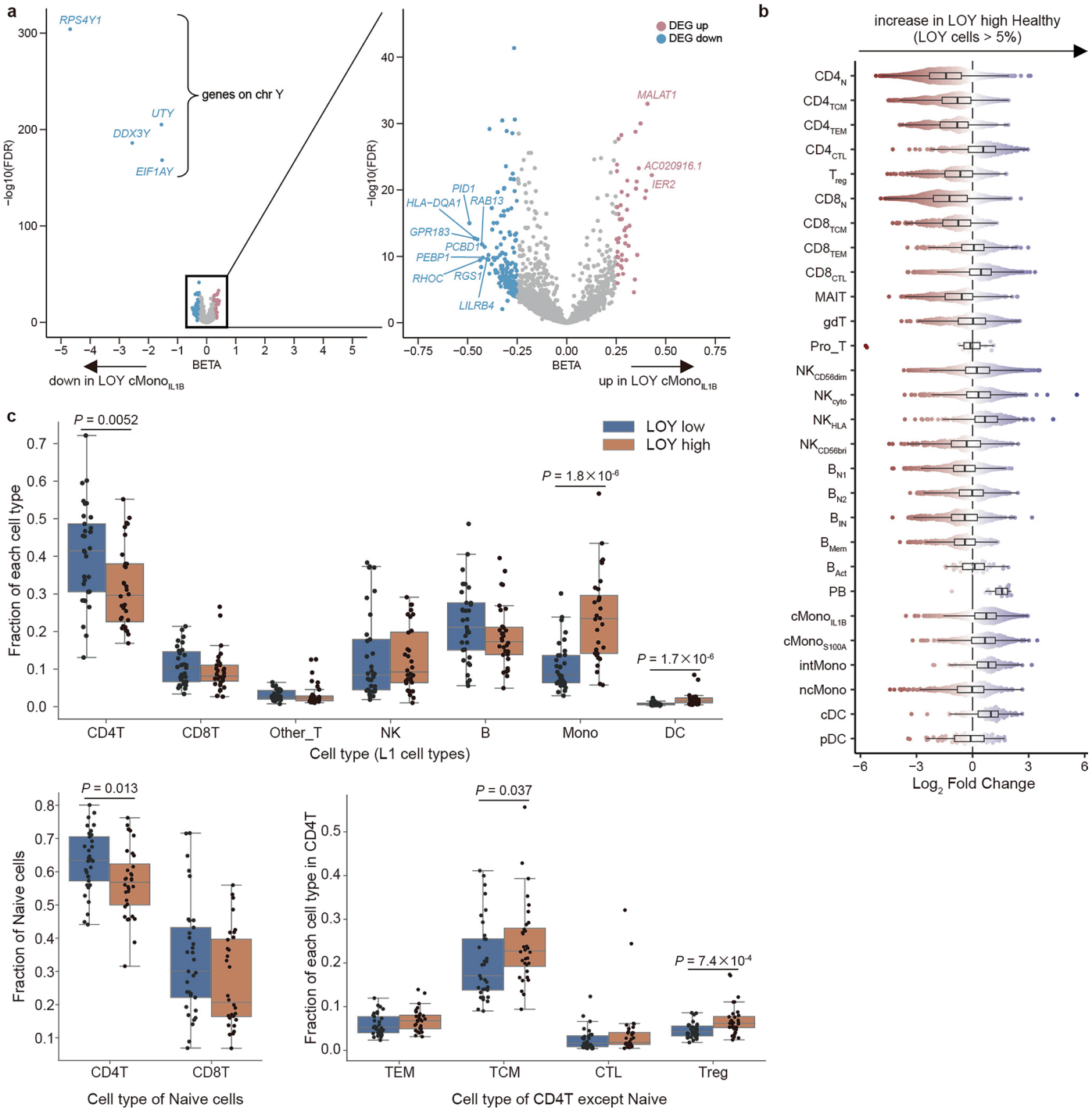
Extended Data Fig. 8 | Characterization of the cells carrying mCAs. (**a**) Fractions of the mutant cells carrying each copy number alterations (CNA) event (left) and those of the mutant cells carrying each copy-neutral loss of heterozygosity (CN-LOH) event colored by setting of sequencing (right). (**b**) Heatmaps showing the in-sample odds ratios of each subcluster (L2) containing the CNA cells in CH01 (top) and CH05 (bottom). (**c**) Differential gene expression (DEG) analysis between the somatically mutant and normal cells for 1p_Loss monocytes (left), 15q_Gain monocytes (middle), and 17q_Gain B cells (right). DEGs are colored in light blue (downregulated) or pink (upregulated), and DEGs on the corresponding chromosomal regions are colored in navy or red. DEGs were significant if they satisfied FDR (adjusted *P* values via the Benjamini-Hochberg method) < 0.05 and log2 fold change > 0.25 (in CH01) or 0.5 (in CH05). (**d**-**g**) Top ten enriched biological pathways of the downregulated DEGs in 1p_Loss monocytes (**d**), the downregulated DEGs in 15q_Gain monocytes (**e**), the downregulated DEGs in 17q_Gain B cells (**f**), and the upregulated DEGs in 17q_Gain B cells (**g**). Dot color indicates the statistical significance of the enrichment (adjusted *P* values via the Benjamini-Hochberg method), and dot size represents gene count annotated to each term.

https://doi.org/10.1038/s41588-025-02266-3



Extended Data Fig. 9 | BCR clonotype of the somatically mutant cells in CH05. (**a**) Network plots showing the similarity of CDR3 amino acid sequence in BCR heavy and light chain from all the samples ($n = 235$), colored by sample (left), cell type (middle), and isotype (right). The somatically mutated CH05 clonotype was surrounded by a plink dotted line. Clonotype clusters with clonal size ≥ 10 are selected. (**b**) Reactivity of antibodies against SARS-CoV-2 antigens in ELISA. The dots denote the mean and error bars do standard deviation measured in triplicate. S309, anti-SARS-CoV-2 S IgG1; CH05, recombinant antibody derived from the CH05 BCR clone; nCoV396, anti-SARS-CoV-2 N IgG1; 23B12, anti-*Candida albicans* IgG1.

https://doi.org/10.1038/s41588-025-02266-3



Extended Data Fig. 10 | Characterization of the LOY cells. (**a**) Differential expression gene (DEG) analysis between the loss of the Y chromosome (LOY) and normal cMono$_{IL1B}$. DEGs were significant if they satisfied FDR (adjusted *P* values via the Benjamini-Hochberg method) < 0.01 and beta > 0.25 and colored in light blue (downregulated) or pink (upregulated). (**b**) Beeswarm plot and box plot describing the distribution of adjusted $\log_2$ fold change between the LOY high and LOY low healthy controls in neighborhoods from 28 cell types (L2). Nodes are neighborhoods, colored by their $\log_2$ fold change adjusted by age and sequencing group. (**c**) Box plots showing the proportions of each cell type in the LOY high and LOY low COVID-19 patients. Wilcoxon two-sided uncorrected *P* values are shown if they satisfied < 0.05. The boxes denote the interquartile range (IQR), the median is shown as the vertical bars, and the whiskers extend to 1.5 times the IQR in (**b**,**c**).

nature portfolio

Corresponding author(s): Yukinori Okada, Atsushi Kumanogoh, Ryuya Edahiro

Last updated by author(s): May 5, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used in data collection.
Data analysis	All data was analyzed with open-source software available online and detailed in the Methods: Cell Ranger (v5.0.0), STAR (v2.7.2a), Scrublet (v0.2.1), scds (v1.10.0), Scanpy (v1.9.1), Seurat (v4.1.0), harmony (v0.1), BWA-MEM (v0.7.17), GATK (v4.2.6.1), SHAPEIT4 (v4.2.1), PLINK(v1.9, v2.00), scran (v1.18.5), PCAForQTL (v0.1.0), tensorQTL (v1.0.7), mashr(v0.2.50), bcl2fastq (v2.19), PRINSEQ-lite (v0.20.4), Trimmomatic (v0.39), bowtie2 (v2.3.5), BMTagger (v3.101), Milo (v1.2.0), Scirpy (v0.10.0), Deep*HLA, DESeq2 (v1.30.1), SHazaM(v1.1.0), Clusterprofiler (v3.14.3), lme4 (v1.1.32), coloc (v5.2.3), PRS-CSx, edgeR (v3.32.0), OlinkAnalyze(v3.4.1), MoChA(2022-05-18), Numbat (v1.3.0), cellSNP-lite (v1.2.2), vireoSNP (v0.5.8), MAST(v1.20.0).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Raw sequencing data of scRNA-seq and the protein expression matrix are available at the Japanese Genotype-phenotype Archive (JGA) with accession codes JGAS000783 (https://ddbj.nig.ac.jp/search/entry/jga-study/JGAS000783)/JGAD000925 (https://ddbj.nig.ac.jp/search/entry/jga-dataset/JGAD000925). A part of the raw scRNA-seq data (nCOVID-19 = 73, nControl = 75) has already been deposited and is available under controlled access at JGA with accession codes JGAS000593/JGAS000543/JGAD000662/JGAD000722. All the raw sequencing data of scRNA-seq can also be accessed through application at the NBDC with the accession code hum0197 (https://humandbs.biosciencedbc.jp/en/hum0197-latest). Genotype data of the subjects are available at European Genome-Phenome Archive (EGA) with the accession code EGAS00001008016 (https://ega-archive.org/studies/EGAS00001008016). The sc-eQTL summary statistics are available at NBDC accession ID hum0197 (https://humandbs.biosciencedbc.jp/en/hum0197-latest), and also available at an interactive browser: https://japan-omics.jp/. OneK1K genotype data were obtained from Gene Expression Omnibus (GSE196830). The processed OneK1K scRNA-seq data (h5ad file) was obtained from Human Cell Atlas (HCA) (https://explore.data.humancellatlas.org/projects/f2078d5f-2e7d-4844-8552-f7c41a231e52). The eQTL summary statistics of OneK1K were obtained from https://onek1k.org/.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	We estimated genotype-based sex for all 235 samples. Out of these, 149 were male, and 86 were female, which were consistent with self-reported gender.
Reporting on race, ethnicity, or other socially relevant groupings	We estimated all 235 samples were Japanese, based on the principal component analysis with the HapMap project samples and the Biobank Japan project samples using the genotyping array data.
Population characteristics	COVID-19 patients (n = 88) are Japanese, the mean age was 64, 73.9% were male, and all of them were tested positive for PCR test. COVID-19 patients were further categorized into groups of moderate (n = 9) and severe (n = 79) according to disease severity based on WHO guidelines. Healthy controls (n = 147) are of Japanese, the mean age was 37, 57.1% were male. Detailed cohort demographics are available in Supplementary Table 1.
Recruitment	We recruited the hospitalized cases diagnosed as COVID-19 by physicians using the clinical manifestation and PCR test results, who were recruited from July 2020 to February 2022 at Osaka University. We included all COVID-19 cases treated at Osaka University who gave consent for the study. Thus, there is little selection bias. All healthy subjects were recruited at Osaka University.
Ethics oversight	This study was approved by the ethical committee of Osaka University Graduate School of Medicine.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample size was not predetermined and all samples that were available were processed. We recruited 88 COVID-19 cases and 147 healthy controls.
Data exclusions	Low quality cells: Cells that had fewer than 1st percentile of UMIs or greater than 99th percentile of UMIs in each sample, as well as cells that contained greater than 10% of reads from mitochondrial genes or Hemoglobin genes, were considered low quality and removed from further analysis. Doublets: Cells labeled as doublets by Scrublet or scds were excluded from the analysis.

	Red blood cells and platelets: Cells manually annotated as red blood cells or platelets were excluded from the analysis.
Replication	sc-eQTL of our cohort were compared with previously published bulk datasets in Japanese (i.e., ImmuNexUT) and sc-eQTL resources in European (i.e., OneK1K) for each corresponding cell type independently.
Randomization	We investigated two major groups; "COVID-19 patients" and "Healthy subjects". COVID-19 patients were further categorized into groups of moderate and severe according to disease severity based on WHO guidelines. Randomization was not applicable to this study.
Blinding	We did not apply blinding of the samples because no intervention was conducted in our study and our study was not case-control one.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	The following antibodies were used in ELISA: - Recombinant monoclonal antibody reconstituted from the BCR of a 17q gain B cell clone from a COVID-19 patient (in this study, detailed in Methods) - anti-SARS-CoV-2 S protein, clone S309 (Pinto et al., Nature 2020, doi:10.1038/s41586-020-2349-y) - anti-SARS-CoV-2 N protein, clone nCoV396 (Kang et al., Nat Commun. 2021, doi:10.1038/s41467-021-23036-9) - anti-Candida albicans, clone 23B12 (Takeda et al., J Allergy Clin Immunol. 2019, doi: 10.1016/j.jaci.2018.07.006) - Goat anti-human IgG-HRP (Southern Biotech, cat. no. 2040-05)
Validation	The high purify of in-house recombinant antibodies was validated by SDS-PAGE. Monoclonal antibodies, S309, nCoV396 and 23B12 were validated by previous publications for specificity. Goat anti-human IgG-HRP antibody (Southern Biotech, cat. no. 2040-05) was validated by the manufacturers for specificity by each production lot.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	Expi293F (Thermo Fisher)
Authentication	None.
Mycoplasma contamination	Cell lines were negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	None.

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*