Title

Constrained Open Chromatin Regions Reveal Functional Elements Shaping Human Traits

Kohei Tomizuka[1]†, Masaru Koido[1,2]†,*, Akari Suzuki[3]†, Soichiro Yoshino[1], Nao Tanaka[1], Yuki Ishikawa[1], Xiaoxi Liu[1], Satoshi Koyama[4], Kazuyoshi Ishigaki[5], Yasuhiro Murakawa[6], Immune transcript/enhancer consortium (ITEC)[#], Kazuhiko Yamamoto[3,7], Chikashi Terao[1,8,9,*]

Affiliations

[1]Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa, 230-0045, Japan.

[2]Laboratory of Complex Trait Genomics, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, 4-6-1, Shirokane-dai, Minato-ku, Tokyo, 108-8639, Japan.

[3]Laboratory for Autoimmune Diseases, RIKEN Center for Integrative Medical Sciences, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa, 230-0045, Japan.

[4]Laboratory for Cardiovascular Genomics and Informatics, RIKEN Center for Integrative Medical Sciences, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa, 230-0045, Japan.

[5]Laboratory for Human Immunogenetics, RIKEN Center for Integrative Medical Sciences, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa, 230-0045, Japan.

[6]RIKEN-IFOM Joint Laboratory for Cancer Genomics, RIKEN Center for Integrative Medical Sciences, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa, 230-0045, Japan.

[7]Drug Discovery Antibody Platform Unit, RIKEN Center for Integrative Medical Sciences, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa, 230-0045, Japan.

[8]Clinical Research Center, Shizuoka General Hospital, 4 Chome-27-1 Kitaando, Aoi Ward, Shizuoka, 420-8527, Japan.

[9]Department of Applied Genetics, The School of Pharmaceutical Sciences, University of Shizuoka, 52-1 Yada, Suruga Ward, Shizuoka, 422-8526, Japan.

[#]Immune transcript/enhancer consortium (ITEC) members

Matteo Guerrini, Hiroaki Hatano, Keiko Hikino, Kazuyoshi Ishigaki, Yuki Ishikawa, Michihiro Kono, Xiaoxi Liu, Yasuhiro Murakawa, Masahiro Nakano, Akiko Oguchi, Raku Son, Akari Suzuki, Kazuhiro Takeuchi, Nao Tanaka, Chikashi Terao, Kohei Tomizuka, Kazuhiko Yamamoto, Soichiro Yoshino

†These authors contributed equally: Kohei Tomizuka, Masaru Koido, Akari Suzuki

*Corresponding authors:

Chikashi Terao (chikashi.terao@riken.jp)

Masaru Koido (masaru.koido@a.riken.jp)

Abstract

Open chromatin regions (OCRs) define cell-type-specific regulatory elements across the genome, yet their functional significance varies, making it challenging to pinpoint biologically essential regions. Here, we introduce CAMBUS (Chromatin Accessibility Mutation Burden Score), a machine-learning framework that identifies active and evolutionarily constrained OCRs by leveraging surrounding DNA sequences. Applying CAMBUS to 29 immune cell types, we identified 66,043 constrained OCRs, which were substantially enriched in the known constraint genome (odds ratio=11.45 (95% confidence

interval 9.33–14.05), $P=4.7 \times 10^{-68}$), while 90% of these OCRs were not prioritized by existing constraint metrics. These OCRs were highly enriched for known enhancers and super-enhancers, independent of known epigenetic markers and annotated regions, and overlapped with regulatory elements implicated in immune-mediated diseases and experimentally validated functional variants, including rare variants, particularly in leukocyte-related traits. Furthermore, CAMBUS revealed cell-type-specific transcriptional regulatory landscapes, linking genetic constraint with gene regulation in immune cells and identifying plausible connections between 1,533 causal variants and 70 complex traits. By defining biologically constrained regulatory elements at high resolution, CAMBUS provides a framework for understanding the selective pressures shaping the non-coding genome and its role in human health and disease.

Introduction

Genome sequences define species identities and evolutionary trajectories by shaping cellular, organ, and functional complexities (O'Bleness, Searles et al. 2012, Pollen, Kilik et al. 2023). Given that protein-coding regions account for only a small fraction of vertebrate genomes (Leypold and Speicher 2021), understanding the role of non-coding regions—particularly active regulatory elements—is critical for deciphering the molecular basis of species identity and function (Ward and Kellis 2012, Franchini and Pollard 2017). Active regulatory regions, such as enhancers, regulate gene expression in a cell-type- and state-specific manner (Andersson, Gebhard et al. 2014, Soskic, Cano-Gamez et al. 2019, Gupta, Weinand et al. 2023, Oguchi, Suzuki et al. 2024).

Several genome-wide experimental methods, including chromatin immunoprecipitation followed by sequencing (ChIP-seq) for histone modifications, DNase I

hypersensitive site sequencing (DNase-seq), and assay for transposase-accessible chromatin using sequencing (ATAC-seq), have been developed to identify candidate regulatory regions. Among them, ATAC-seq has become widely adopted due to its technical simplicity and robust performance (Buenrostro, Giresi et al. 2013, Yan, Powell et al. 2020). Compared with ChIP-seq peaks, ATAC-seq-defined open chromatin regions (OCRs) provide high positional specificity for identifying regulatory elements (Consortium, Moore et al. 2020). Because chromatin accessibility is a prerequisite for gene transcription, mapping and characterizing OCRs—particularly those that are cell-type-specific or conserved across multiple lineages—offers crucial insights into general and context-specific regulatory mechanisms, evolutionary constraint, and the basics of complex traits (Boyle, Davis et al. 2008, Zhang, Hocker et al. 2021).

Although genome-wide association studies (GWAS) in human have highlighted the enrichment of disease-associated variants within non-coding regions (Maurano, Humbert et al. 2012, Claussnitzer, Cho et al. 2020), and particularly within regulatory elements (Trynka, Sandor et al. 2013, Soskic, Cano-Gamez et al. 2019, Zeng, Bendl et al. 2024), distinguishing active regulatory elements from non-active OCRs remains challenging (Maurano, Humbert et al. 2012, Visscher, Wray et al. 2017, Schaid, Chen et al. 2018), as ATAC-seq measures chromatin openness but does not directly assess regulatory activity (Thibodeau, Uyar et al. 2018, Klemm, Shipony et al. 2019). Therefore, integrating additional layers of information is necessary to systemically predict active and functional regulatory regions.

Constrained genomic regions, which show evidence of purifying selection, serve as important markers of functionally active elements (Ward and Kellis 2012, Radke, Sul et al. 2021). The systematic identification of constrained, cell-type-specific active regulatory regions across diverse tissues remains an open challenge. Recent large-scale human whole-genome sequencing efforts have enabled the development of constraint metrics

such as Gnocchi (Chen, Francioli et al. 2024), which leverage local mutability to infer constrained sites. Consortia efforts such as ENCODE Phase 3 have combined chromatin accessibility with histone modification ChIP-seq signals to annotate candidate cis-regulatory elements (cCREs), but cCREs cannot fully capture functional regulatory regions.

In addition to evolutionary constraint and epigenomic signals, chromatin accessibility quantitative trait loci (caQTLs), which identify genetic variants that alter chromatin openness, provide further evidence that genetic variation at regulatory regions influences complex traits (Kumasaka, Knights et al. 2016, Gate, Cheng et al. 2018). However, caQTL mapping requires large cohorts and is limited in its ability to assess rare variants. Recently, machine learning models trained on DNA sequences have emerged as a powerful alternative, capable of predicting regulatory activity and inferring variant effects without requiring population-level data (Zhou, Theesfeld et al. 2018, Eraslan, Avsec et al. 2019, Koido, Hon et al. 2023). These models have enabled the tissue- and cell-type-specific prediction of gene expression and chromatin features from sequence alone (Zhou and Troyanskaya 2015, Koido, Hon et al. 2023).

Together, these observations highlight the need for methods that integrate evolutionary constraint, chromatin accessibility, and DNA sequence features to prioritize functional regulatory elements. Motivated by this gap, we developed CAMBUS (Chromatin Accessibility Mutation Burden Score), a new framework for prioritizing active and evolutionarily constrained OCRs. By leveraging sequence-embedded constraint information (Fig. 1), CAMBUS provides a principled approach to identify regulatory elements with likely functional importance across diverse cell types. We assessed the characteristics of CAMBUS and their cell-type specificity using OCRs in various cell types in peripheral blood mononuclear cells (PBMCs). By leveraging CAMBUS, we identified 1,533 cell-type-specific regulatory variants, including rare variants, with potential functional significance in complex traits across European (EUR) and East Asian (EAS) ancestries (Fig. 1D and 1E). We

demonstrated how integrating genetic constraint with chromatin accessibility can refine the biological interpretation of non-coding regulatory variants and their roles in transcriptional regulation in a cell-type-specific manner.

Results

Comprehensive ATAC-seq profiling of 29 immune cell types

We sorted PBMCs (peripheral blood mononuclear cells) and Neutrophils from 55 Japanese individuals (Methods). PBMCs were sorted into 27 immune cell types by cell sorter. We conducted ATAC-seq for PBMC, Neutrophil, and the 27 cell types (hereafter, 29 cell types) in eight subgroups (CD4; $CD4^+$ T-cells, CD8; $CD8^+$ T-cells, B; B cells, Mono; Monocytes, DC; dendritic cells, NK; Natural Killer cells, Neutrophils, and PBMC) (Supplementary Table 1). The number of cell types in this study is comparable to that in recent scATAC-seq studies using PBMC (Zhang, Wang et al. 2020, Benaglio, Newsome et al. 2023). Notably, the 29 cell types include minor cell types, such as double-negative B-cells (~5% in B-cells) (Li, Li et al. 2021). After stringent quality control, we obtained 7,939,714 OCRs that covered 498 Mb (16.1 %) of the human genome (Methods). The median number and size of OCRs were 252,668 OCRs and 276 bp, respectively. Among the 7,939,714 OCRs, 330,367 (4.2%) were observed in only one cell type, and 437,368 (5.5%) were observed in only one subgroup. Both hierarchical clustering and dimension reduction using principal component analysis (PCA) and uniform manifold approximation and projection (UMAP) showed that these OCRs could clearly distinguish the eight subgroups but not all 29 immune cell types (Extended Data Fig. 1).

Detecting constrained OCRs by CAMBUS

To infer constraint levels in the OCRs, we first developed machine learning models, CAMBUS ML (Chromatin Accessibility Mutation Burden Score machine learning), to predict the OCRs of each of 29 cell types from DNA sequence patterns, and calculated *in silico* saturated mutation effects- quantifying the predicted changes in accessibility probability for all possible single nucleotide mutations within 2kb of each OCR (whether or not the mutations were reported) (Methods). We aggregated and took the mean of mutation effects within 2kb of each OCR and called this value CAMBUS. Hereafter, we defined OCRs with negative CAMBUS as CAMBUS+ (where the '+' denotes positive constraint) based on the local false discovery rate (FDR) method (Fig. 2A, 2B; Method). Besides, we defined relaxed CAMBUS+ OCRs, a more relaxed CAMBUS+ set by extracting the bottom 5% of the overall CAMBUS distribution. The median number of CAMBUS+ OCRs was 1,584, covering 0.056% of the human genome, while that of relaxed CAMBUS+ OCRs was 12,632, covering 0.33%. Notably, read counts from ATAC-seq were not a determinant factor for CAMBUS+ OCRs (Extended Data Fig. 2A), reinforcing that our machine-learning approach incorporated information from DNA sequence patterns into OCRs, not from technical fluctuations or noises.

To confirm our key assumption that CAMBUS+ OCRs (or relaxed CAMBUS+ OCRs) are under positive constraint, we assessed whether CAMBUS+ OCRs overlapped with orthogonal seven constrained metrics (Methods). As expected, five out of the seven constraint metrics supported the constraint hypothesis for CAMBUS+ OCRs across 29 cell types (Method; Supplementary Table 3). Notably, we found that CAMBUS+ OCRs significantly cover the strongly constrained regions (one-sided Wilcoxon-test $P = 2.5 \times 10^{-69}$ for CAMBUS+ OCRs in NaiveCD4). Although CAMBUS+ OCRs contained a greater proportion of coding regions than control OCRs (Extended Data Fig. 2B), the same trend of enrichment was observed even when the analysis was restricted to non-coding regions

(one-sided Wilcoxon-test P = 9.1 × 10^{-46} for CAMBUS+ OCRs in NaiveCD4; Fig. 2C; Extended Data Fig. 2C). Of note, when we analyzed each of CAMBUS+ OCRs, 59,605 (90%) of CAMBUS+ OCRs were not prioritized by existing constraint metrics (Methods). Nevertheless, the CAMBUS+ OCRs showed the strongest constraint defined by Gnocchi (odds ratio (OR)=11.45 (95% confidence interval (CI) 9.33–14.05), P=4.7 × 10^{-68} by Fisher's Exact test) among the compared regulatory elements (Fig. 2D; Supplementary Table 4). Meta-analysis demonstrated the consistent tendency across the 29 cell types (OR=8.71 (95% CI 7.85–9.66), P=3.1 × 10^{-365} from random-effect meta-analysis; Fig. 2E), outperforming the ENCODE promoter-like regulatory elements (OR=7.3), the strongest constraint candidate regulatory elements in the previous study (Chen, Francioli et al. 2024). Collectively, these indicate that CAMBUS+ OCRs could uniquely find regions under very strong positive constraints. The relaxed CAMBUS+ OCRs also showed the same patterns with slightly reduced magnitude (Extended Data Fig. 2D).

Previous studies (Silvert, Quintana-Murci et al. 2019) demonstrated that enhancers, especially activated in T-cells, are enriched in archaic genome sequences due to positive selection. Consistently, we observed that CAMBUS+ OCRs showed significant enrichment in the selection signals (Nait Saada, Kalantzis et al. 2020) (OR=1.4 (95% CI 1.3–1.5), P=1.1 × 10^{-26} from random-effect meta-analysis; Extended Data Fig. 2E), the Neanderthal genome sequences observed in the Japanese population (Liu, Koyama et al. 2024) (OR=1.4 (95% CI 1.4–1.5), P=4.8 x 10^{-32} from random-effect meta-analysis; Extended Data Fig. 2F), and the Neanderthal and Denisovans' sequences observed in 1000 Genomes populations (Supplementary Table 5) (Browning, Browning et al. 2018). Besides, we observed the enrichment of CAMBUS+ in human ancestor quickly evolved regions (HAQER) (Mangan, Alsina et al. 2022) (OR=2.35 (95% CI 1.91–2.89), P=8.3 × 10^{-16}).

Taken together, CAMBUS+ OCRs can define novel constrained regions, well supported by other constraint-relevant metrics.

Enrichment of active regulatory elements by selecting CAMBUS+ OCRs

As expected, the OCRs obtained in this study were well overlapped with the known candidate regulatory elements; for example, 73.66% of all OCRs in NaiveCD4 were overlapped with super-enhancers (Jiang, Qian et al. 2019) (the median of 29 cell type is 74.70%) and 4.49% with FANTOM5 enhancers (Andersson, Gebhard et al. 2014) (the median of 29 cell type is 5.10%; Supplementary Table 7). Then we explored whether CAMBUS+ OCRs could aid in prioritizing functional regulatory elements in comparison with control OCRs (Methods). As a result, CAMBUS+ OCRs significantly enriched for the known regulatory elements, especially active enhancers, like super-enhancers (NaiveCD4: OR=9.07 (95% CI 6.80–12.11), Fisher's exact test $P = 2.0 \times 10^{-102}$) and FANTOM5 enhancers (NaiveCD4: OR=6.16 (95% CI 5.39–7.03), Fisher's exact test $P = 6.3 \times 10^{-111}$) (Supplementary Table 8). Meta-analysis showed a consistent trend across 29 cell types (OR=6.63 (95% CI 5.50–7.99), $P = 1.1 \times 10^{-87}$ from random-effect meta-analysis for super-enhancers; Extended Data Fig. 2H, and OR=5.19 (95% CI 4.87–5.54), $P = 4.9 \times 10^{-547}$ from random-effect meta-analysis for FANTOM5 enhancers; Extended Data Fig. 2I). Remarkably, after adjusting for known constraint metrics, we still observed statistically significant enrichment for regulatory elements in CAMBUS+ OCRs across the 29 cell types (Fig. 2F; Supplementary Table 9). Reflecting the characteristics of the analyzed cell types (blood and immune cell types), a comparison of enrichment across four tissues (blood/hematopoietic, brain, kidney, and lung) revealed the strongest prioritization of blood's regulatory elements (CD4 as a representative: super-enhancers; OR=6.83 (95% CI 6.07–7.71), $P=1.1 \times 10^{-217}$; FANTOM5 enhancers; OR=7.64 (95% CI 6.64–8.78), $P=8.7 \times 10^{-179}$ from logistic regression; Fig. 2G; Supplementary Table 10). This pattern was

consistently observed across 29 cell types (Supplementary Table 11). We also confirmed that the relaxed CAMBUS+ OCRs exhibited the same tendency (Extended Data Fig. 2J; Supplementary Table 12). Taken together, CAMBUS+ OCRs (and relaxed CAMBUS+ OCRs) effectively capture active regulatory elements under positive constraints in a cell-type-specific manner, particularly enhancers, that were not well captured by the previous methods.

Heritability enrichment of CAMBUS+ in blood-relevant diseases and traits

Since the majority of trait heritability is captured by regulatory regions, we computed heritability enrichments in CAMBUS+ OCRs using stratified LD score regression (S-LDSC) (Finucane, Bulik-Sullivan et al. 2015) conditioning on known functional annotations and all OCRs from this study (Methods). We used GWAS summary statistics for 33 complex traits, including 14 immune-mediated traits (e.g., multiple sclerosis (MS), rheumatoid arthritis (RA), and Atopic dermatitis (AD)), from European (EUR) and East Asian (EAS) populations (Supplementary Table 13).

As a result, CAMBUS+ OCRs demonstrated the highly significant heritability enrichment in a wide variety of immune-mediated diseases (FDR < 5%; Fig. 3A; Supplementary Table 14). Particularly, we found plausible pairs of immune traits and cell types supported by the strong enrichment of heritability: MS and T_h17 (Moser, Akgun et al. 2020), and systemic lupus erythematosus (SLE) and B cells (Jarvinen, Hellquist et al. 2012, Fillatreau, Manfroi et al. 2021) (Extended Data Fig. 3A). Other plausible pairs included pairs of RA with CD4 or $CD8^+$ T cells, and AD with CD4 (particularly Th2 (Honda and Kabashima 2020)), or $CD8^+$ T cells. Remarkably, these enrichments were not observed in other constraint metrics (Fig. 3B).

Based on these results, we hypothesized that polygenic risk scores (PRS) prioritized by CAMBUS+ OCRs could enhance their predictive ability. To evaluate this hypothesis, we

focused on CAMBUS+ OCRs in double-negative B cells (DN B), which showed significant enrichment for the heritability of SLE in EAS ancestry (Extended Data Fig. 3A) and was also known for its functional associations with SLE (Chung, Gong et al. 2023). We constructed the PRS based on variants overlapping CAMBUS+ OCRs and evaluated the additional performance of the conventional PRS by five-fold cross-validation using 137,620 individuals (Methods). A fixed effect meta-analysis demonstrated that the additional variants significantly improved the predictive performance of PRS ($P = 3.12 \times 10^{-6}$; Supplementary Table 15; Methods). These suggest that CAMBUS+ OCRs help us to understand the basics of the complex traits in a polygenic manner and in a cell-type specific manner and help us to construct better PRS by prioritizing target cell types.

Putative causal variants were enriched in CAMBUS+ OCRs

To further search the potential usefulness of CAMBUS+ OCRs in a single variant resolution, we assessed whether CAMBUS+ OCRs could prioritize or pinpoint the putative causal variants for complex traits. As expected, the median regional length of CAMBUS+ (845 bp) was much shorter than that of 95% credible sets (3.0M bp for EUR ancestry and 1.1M bp for EAS ancestry), which may potentially pinpoint a variant. We observed CAMBUS+ OCRs showed significant enrichment in the pathogenic regulatory variants cataloged in the Human Gene Mutation Database (HGMD) (Stenson, Ball et al. 2003) (OR=2.70 (95% CI 2.43–2.99), $P = 1.7 \times 10^{-77}$; from random-effect meta-analysis; Extended Data Fig. 2G; Supplementary Table 6). This is in line with overlap between CAMBUS+ OCRs and super-enhancers or FANTOM5 enhancers, both of which are known to be enriched for causal variants (Hnisz, Abraham et al. 2013). We further analyzed the large-scale statistical fine-mapping studies in 94 traits in EUR ancestry (Kanai, Ulirsch et al. 2021) and 62 traits in EAS ancestry (Koyama, Liu et al. 2024). The CAMBUS+ OCRs were significantly enriched for the putative causal variants in 95% credible sets (median OR:3.85

and 3.94 in EUR and EAS, respectively, Extended Data Fig. 4A). In line with a data source of OCRs, CAMBUS+ OCRs exhibited significantly higher enrichment for the putative causal variants of hematopoietic traits, especially leukocyte-related traits (Methods, Supplementary Table 21) than non-hematopoietic traits (two-sided *t*-test P = 1.0×10^{-11} in EUR ancestry and 1.4×10^{-2} in EAS ancestry for leukocyte-related traits; Extended Data Fig. 4B). As an example, CAMBUS+ OCRs in type 1 helper T-cells demonstrated stronger enrichment for variants with high posterior inclusion probability (PIP) (>0.9) for hematopoietic traits in EAS ancestry (OR=7.62 (95% CI 3.60–16.18), P = 1.2×10^{-7} from logistic regression; Fig. 4A), and this enrichment was further increased when restricted to the leukocyte-related traits (OR=11.0 (95% CI 4.04–29.96), P = 2.7×10^{-6}; Fig. 4A). Meta-analysis showed that this trend of leukocyte-related traits was consistent across 29 cell types (OR=8.99 (95% CI 7.85–10.29), P = 8.8×10^{-221} for EUR ancestry; OR=19.24 (95% CI 16.03 – 23.09), P = 3.9×10^{-221} for EAS ancestry; Fig. 4B). Taken together, CAMBUS+ OCRs can pinpoint or at least narrow down causal variants.

Notably, even for rare variants (minor allele frequency (MAF) < 0.01), the putative causal variants of hematopoietic traits were significantly enriched in CAMBUS+ OCRs (Methods; OR=4.68 (95% CI 2.66–7.99), two-sided Fisher's exact test P = 2.23×10^{-6}; Fig. 4C). This suggested that CAMBUS+ OCRs can potentially prioritize even rare causal variants which are difficult to prioritize in the conventional statistical fine-mapping (Schaid, Chen et al. 2018).

We found plausible examples of putative causal variants that were not covered by known constraint regions (Gnocchi) but were identified by CAMBUS+ OCRs, leading to the biological interpretation of GWAS findings. First, rs13122305 was the only one overlapping with CAMBUS+ OCRs in monocytes among seven candidate causal variants in a credible set for monocyte count (Fig. 4D). The nearest gene of this variant is *COQ2*, whose anti-inflammatory functions via regulating gene expression were reported (Armanfar, Jafari et al.

2015, Mantle, Heaton et al. 2021). Additionally, we found that the Activity-by-Contact (ABC) model's prediction (ABC enhancer) in monocytes only overlapped with rs13122305 and that *COQ2* was also supported by the ABC model with the second-highest ABC score (0.025) among the eight target genes. Second, among the three candidate causal variants in a credible set in the FCGR region for Albumin/Globulin ratio (AG), only rs2487451 overlapped with CAMBUS+ OCR. Non-classical monocytes, known to display antigen presenting properties (Mukherjee, Kanti Barman et al. 2015), were the only cell type whose OCR overlapped with the candidate causal variant. The nearest gene was *FCGR3B*, encoding a receptor (FcγRIIIB) that binds to the Fc region of IgG (Immunoglobulin) antibodies, which is known to be expressed in non-classical monocytes (Nassir, Tambi et al. 2021). Third, rs55646091, a putative causal variant for asthma (PIP = 0.99), overlapped with CAMBUS+ OCR of regulatory T-cells (T-reg) (Fig. 4E). The nearest gene of this variant was *LRRC32* (leucine-rich repeat-containing 32), which regulates TGF-beta. LRRC32 was potentially involved in asthma (Li, Wang et al. 2023) and whose encoding protein was known as a surface marker for T-reg. Furthermore, this variant-gene pair was supported by the ABC model and *LRRC32* displayed the highest ABC score (ABC score = 0.041 that is much higher than the threshold corresponding to 70% recall and 59% precision) (Fulco, Nasser et al. 2019) among the nine genes connected with the regulatory element overlapping with the OCR and the variant. Fourth, we found that rs8009224 was the only one that overlapped with CAMBUS+ OCRs of plasmablast (Extended Data Fig. 5A) among the seven putative causal variants for hemoglobin A1c levels. This is an intronic variant of *MIDEAS* (Mitotic Deacetylase Associated SANT Domain Protein), which is low frequency in both 1KG EUR (MAF = 3.4%) and EAS (MAF = 1.2%) and also associated with mean corpuscular hemoglobin concentration (Barton, Sherman et al. 2021). The overlapping regulatory element was connected to the 16 genes by the ABC enhancer in B-cell, and the putative target gene with the second highest ABC score is *MIDEAS* (ABC

score = 0.097, much higher than the proposed threshold corresponding to 70% recall and 59% precision). Lastly, we demonstrated that CAMBUS+ in T-reg could prioritize only one putative causal variant (rs9814991) from numerous variants (fifty-seven) in the credible set for eosinophil count (Extended Data Fig. 5B). These indicate that CAMBUS+ OCRs, together with statistical fine-mapping, could pinpoint putative causal variants of complex traits in a single base resolution and in a cell type-specific manner.

Discussion

In this study, we developed CAMBUS based on the hypothesis that DNA sequencing patterns surrounding OCRs contain hidden information defining OCR constraints. Although CAMBUS+ OCRs were derived from a fundamentally different approach, they were well supported by conventional constraint metrics as expected, while also capturing high-resolution, cell-type-specific regions that are both active and constrained, beyond existing annotations. An alternative potential approach using ATAC-seq (or DNase-seq) could prioritize regions with high accessibility (i.e., regions with many read counts). However, previous studies have suggested no relationship between accessibility and functionality or constraints (Roadmap Epigenomics, Kundaje et al. 2015, Kuderna, Ulirsch et al. 2024, Martens, Fischer et al. 2024). Consistently, accessibility was not a determinant factor in defining CAMBUS+ OCRs. Thus, leveraging surrounding DNA sequence patterns through machine learning offers a unique and effective strategy for identifying biologically important OCRs.

CAMBUS+ OCRs can mark active and functional regulatory elements, particularly tissue-specific active enhancers, even after being adjusted for known constraint metrics (Gnocchi, phyloP46way, phyloP100way, cactus241way, phyloPPrimates, GERP and DR). Furthermore, CAMBUS+ OCRs captured a significant proportion of the heritability for many

complex traits, highlighting their biological and genetic relevance. CAMBUS+ OCRs effectively prioritized regulatory causal variants for immune-related traits, even rare causal variants. By defining biologically constrained regulatory elements, CAMBUS+ OCRs will provide a foundation for understanding the regulatory architecture of complex traits and their biological insights into genetic variations especially in non-coding regions.

We demonstrate that CAMBUS+ OCRs can effectively prioritize causal rare variants for complex traits. A previous study showed that rare variants tend to be enriched in the constraint genome (Karczewski, Francioli et al. 2020), consistent with our study as CAMBUS+ OCRs can mark such constrained regions. In GWAS using SNP arrays, low statistical power and low imputation accuracy for rare variants make it difficult to detect rare variant associations and identify causal rare variants through statistical fine-mapping. Despite these challenges, CAMBUS+ OCRs can facilitate the rare variant prioritization that might otherwise be missed.

Eosinophilic asthma is a subtype characterized by elevated levels of eosinophils (Nelson, Bush et al. 2020). In this analysis, we identified a variant (rs55646091) associated with asthma linked to increased eosinophil counts (PIP = 0.03). This variant exhibited positive effect sizes in both traits (0.20 for asthma, 0.072 for eosinophil counts), suggesting that carriers of this variant demonstrate elevated eosinophil levels, which contribute to inflammation and the onset of asthma (Harb and Chatila 2023). Notably, this variant overlaps specifically with the OCR in T-reg cells among 29 blood cell types analyzed. The nearest gene of this variant is *LRRC32*, which regulates TGF-beta, a known surface marker of T-reg cells, which has been previously implicated in immune regulation (Tran, Andersson et al. 2009). The findings from this analysis provide additional evidence supporting a biological mechanism in which the variant, through its effects on T-reg cells, leads to increased eosinophil counts, thereby contributing to the development of asthma. This

insight deepens our understanding of the role of T-reg cells and *LRRC32* in asthma pathogenesis, particularly in eosinophilic inflammation.

In this study, we focused on various immune cell types; however, our approach is broadly applicable to other tissues and cell types. Because calculating CAMBUS relies on *in silico* saturated mutagenesis across millions of OCRs, our CNN-based framework is practical for large-scale regulatory annotation across diverse cell types, achieving substantially greater computational efficiency than recent large-scale generative AI models (e.g., Evo 2 (Brixi, Durrant et al. 2025) will require more than 200 years for the task on a single H100 GPU (Supplementary Note)). As GWAS results continue to accumulate, integrating CAMBUS+ OCRs with these expanding resources has the potential to uncover further insights into the regulatory mechanisms underlying complex traits. Moreover, this method can be applicable to OCRs identified at the single-cell level, further extending its applicability. Taken together, CAMBUS+ OCRs not only capture constrained and functional regulatory elements but also facilitate the interpretation of non-coding genetic variants by prioritizing causal variants in a cell-type-specific manner without relying on LD information and allele frequencies.

Method

Sample and library preparation

We collected peripheral blood (PB) cells from 55 healthy Japanese volunteers and sorted them into 29 cell types. PBMCs were isolated by density gradient separation with Ficoll-Paque (GE healthcare) immediately after blood drawing. Erythrocytes were lysed with a potassium ammonium chloride buffer, and non-specific binding was blocked with Fc-

gamma receptor antibodies. PBMC is stained with cell surface specific antibodies (Supplementary Table 22). 27 immune cell populations were sorted by FACS (BDAria III) using the gating strategy (Supplementary Table 1; Extended Data Fig. 6). In addition, neutrophils were isolated by MACSxpress Neutrophil Isolation Kits human (Miltenyi Biotec) from 4 × 106 PBMC cells. Sorted cells, isolated PBMC, and Neutrophils were stored at −80°C in CytoStor-10 (Charles River Laboratories Cell Solutions). Libraries for ATAC-seq were prepared using a modified protocol of Omni-ATAC (Corces, Trevino et al. 2017). Genomic DNA from whole blood was isolated using QIAamp DNA Blood Mini Kits, respectively. ATAC-seq was performed using 150bp paired-end on Illumina NovaSeq6000. In total, we obtained a median of 13.34M (+/- 7.56M) reads per sample. This study was approved by the Ethics Committees of RIKEN. Written informed consent was obtained from each volunteer.

Mapping ATAC-seq data

ATAC-seq data were processed following the ENCODE ATAC-seq pipeline (version 1.8.0) as follows: Raw reads were trimmed using cutadapt (version 2.5) and aligned to human genome assembly GRCh38 analysis set using bowtie2 (version 2.3.4.3) for each sample. After alignment, duplicated reads were removed using MarkDuplicates (Picard version 2.20.7-SNAPSHOT), retaining uniquely mapped reads and improperly paired reads using samtools (version 1.9), and removed reads that mapped to the mitochondrial genome.

In total, we obtained a median of 5.39M (s.d. 3.28M) QCed reads per sample.

ATAC-seq peak calling

For each 29 immune cell type, we used the samtools merge command to merge the QCed reads for each cell type and used MACS2 (version 2.1.1) to identify the ATAC-seq peaks (OCRs) following a recent study (Gate, Cheng et al. 2018) with parameters (-p 0.01 --shift -75 –extsize 150 --nomodel). We obtained a median of 253,553 (±66,738) OCRs with a median peak width of 271bp (±18bp) (FDR<0.05). Using the alignment file for each cell type, we calculated the read counts from the first mates for each OCR with bedtools (version 2.30.0)'s coverage function.

Quality control for ATAC-seq dataset

We first assessed the quality of ATAC-seq data using Pearson correlations of chromatin accessibility among the samples for each cell type, using 614,516 peaks that were identified from across all samples using the same parameters described above (these peaks were only used for quality control purposes). We filtered OCRs out when the Pearson correlations were not greater than 0.65 (determined by manual inspection of the distribution of these correlations). Additionally, we excluded outlier samples with low Pearson correlations by visual inspection (Supplementary Table 16). Samples with a mapped rate of less than 0.5 were further excluded. Finally, we obtained 1,386 samples (the median number of donors per cell type is 48; Supplementary Table 1).

Hierarchical clustering and dimension reduction

We first chose the top 5,000 accessible OCRs (defined by TPM) from each cell type and merged them across 29 cell types, resulting in a total of 145,000 unique regions. We then created a binary matrix (0, 1) to indicate whether each 300bp bin across the genome was overlapped by the corresponding OCRs. Hierarchical clustering was performed using Ward's method of Euclidean distance using the binary matrix. PCA for the binary matrix was performed using the binaryPCA function from the scBFA package (version 1.16.0) (Li and Quon 2019) in R with default parameters. UMAP was performed with the top five PCs using the umap function from the scikit-learn package (version 0.18.1) in Python with default parameters.

Strategy for the CAMBUS ML model

The CAMBUS ML model was built upon the DeepSEA architecture, a deep CNN model (Zhou and Troyanskaya 2015). We used the Selene library (v0.4.7) to train the CAMBUS ML model. A 5kb reference sequence, randomly chosen with Selene, is used as input features, five times longer than the sequence used in the original DeepSEA model to capture broader genomic contexts. The model outputs a sigmoid function-derived score, which can be interpreted as a proxy for the likelihood that the OCRs overlap with the central 200bp of the input sequence (see Supplementary Table 17 for the configuration details). We referred to this predicted probability for chromatin accessibility as the accessibility probability.

We combined 7,939,714 OCRs across 29 cell types using the `bedtools merge` function and obtained 1,349,934 non-overlapping unique input sequences. According to Selene's default settings, we used chromosomes 8 and 9 as testing, chromosomes 6 and 7 as

validation, and the other autosomal chromosomes as training data. The details of Selene's parameters are described in Supplementary Table 17.

In silico saturated mutagenesis using the trained CAMBUS ML model

For all bases within 1kb from the center of each OCR, we used the trained CAMBUS ML model to calculate the predictions for all possible single-based mutations. This comprehensive approach, inspired by the previous study working on transcript expression levels (Zhou, Theesfeld et al. 2018), provides an unbiased analysis of mutational effects across the entire sequence, rather than focusing on only specific known polymorphisms. To quantify the mutation effect, we calculated the difference in log fold changes of predictions between the reference and alternative alleles for each site and averaged them across all bases. In total, we obtained 505 million mutation effects (median across the 29 cell types).

Identification of putative constraint open chromatin region

We modeled the distribution of the mean mutation effect (*in silico* saturated mutagenesis) within 2kb of each OCR (CAMBUS) as a mixture of a Gaussian null distribution. We hypothesized that the OCRs with negative CAMBUS are likely positive constraint regions and OCRs with neither negative nor positive CAMBUS as control OCRs. We calculated local FDR with the locfdr function in the locfdr R package (version 1.1-8) for each cell type. This analysis identified a median of 1,584 putative positive constraint OCRs across 29 cell types with FDR < 0.5, referred to as CAMBUS+ OCRs, and a median of 248,749 control OCRs across 29 cell types (Zhou, Theesfeld et al. 2018). We also defined relaxed CAMBUS+ OCRs by extracting the bottom 5% of the CAMBUS distribution for each cell type (Supplementary Table 19).

Enrichment of CAMBUS+ OCRs in constraint regions

We evaluated whether CAMBUS+ OCRs are enriched in constraint regions of the genome using Gnocchi. We used the 1,984,900 QCed Gnocchi (see URLs) and calculated the OR following the methodology described in the previous paper (Chen, Francioli et al. 2024) for 1,843,559 non-coding 1-kb windows. In brief, we counted the number of overlapping 1kb windows within the following bins: (∞, -4), [−4,−3), [−3,−2), [−2,−1), [−1,−0), [0,1), [1,2), [2,3), [3,4), [4,∞) with each CAMBUS+ OCR. ORs were calculated by comparing these overlaps with all 1kb windows. Fisher's exact test was conducted to estimate the statistical significance (p-value). The OR of the known regulatory elements for each bin was obtained from the gnomAD download site (see URLs) and used for the comparison.

To calculate the overlapping of the constraint metrics in CAMBUS+ OCRs, we used the seven constraint metrics (see URLs): (1) The Gnocchi, estimated by comparing observed allele counts in diverse human populations with the estimated counts in each 1kb region of the human genome (Chen, Francioli et al. 2024), (2) phyloP, defined by phylogenetic P-values (phyloP) scores derived from conservation across various species: phyloP46way (Pollard, Hubisz et al. 2010), phyloP100way, phyloP241way (Zoonomia 2020), based on alignments of 46, 100, and 241 species, respectively, (3) Primate-specific PhastCons (phyloPPrimates), defined by phylogenetic hidden Markov model (phylo-HMM) across 100 primate species (Zoonomia 2020), (4) Depletion Rank (DR), estimated by comparing observed allele counts in EUR populations with the estimated counts in each 500bp region

of the human genome (Halldorsson, Eggertsson et al. 2022), (5) Genomic Evolutionary Rate Profiling (GERP), defined by rejected substitution scores (RS) that quantify evolutionary constraint using alignments of diverse mammalian genomes (Cooper, Stone et al. 2005). The phyloP46way mapped the positions of those datasets to the hg38/GRCh38 coordinate using liftOver as described previously. For the Gnocchi and DR, the overlapping regions were considered to be overlapped if they had more than 50% overlap with each other.

To determine whether CAMBUS+ OCRs were prioritized by existing constraint metrics, we employed the following approach: First, we calculated the mean values for each of the seven constraint metrics overlapping for each CAMBUS+ OCR. Second, a CAMBUS+ OCR was considered not prioritized by existing constraint metrics if all seven mean values failed to meet their recommended thresholds; Gnocchi > 4 (Chen, Francioli et al. 2024), phyloP100way with score > 3.0075 (FDR5%, following the previous approach (Sullivan, Meadows et al. 2023)), phyloPPrimates with score > 0.961 (Sullivan, Meadows et al. 2023), cactus241way > 2.27 (Sullivan, Meadows et al. 2023), DR < 0.01 (Halldorsson, Eggertsson et al. 2022), GERP > 2 (a RS score threshold of 2 provides high sensitivity while still strongly enriching for truly constrained sites, as described in the UCSC Genome Browser documentation at 2025/7/22) (Fu, Liu et al. 2014)). Regarding the phyloP46way, we used top10% score (0.533) as a threshold since even the maximum score (0.655)is not nominally significant (corresponding P value is 0.221) due to lack of power since the number of species is small (Pollard, Hubisz et al. 2010).

Enrichment of CAMBUS+ OCRs in the positive selection signals

We evaluated whether CAMBUS+ OCRs are enriched in the positive selection signals previously reported (Nait Saada, Kalantzis et al. 2020). For each CAMBUS+ OCR, we

counted overlapping annotations for the positive selection signals observed in the EUR population and estimated the OR compared with all OCRs using a logistic regression model. To perform enrichment analysis of HAQERs (Mangan, Alsina et al. 2022), we first created merged OCRs by combining CAMBUS+ OCRs from 29 different cell types (merged CAMBUS+ OCRs). The number of overlapped merged CAMBUS+ OCRs in the HAQERs was counted, and OR was estimated compared with all OCRs using a logistic regression model.

Also, we evaluated whether CAMBUS+ OCRs are enriched in the archaic genome. For each CAMBUS+ OCR, we counted overlapping annotations for the Neanderthal genome sequences observed in the Japanese population (Liu, Koyama et al. 2024) and the Neanderthal and Denisovans' sequences observed in 1000 Genomes populations (Browning, Browning et al. 2018), and estimated the OR compared with all OCRs using a logistic regression model.

Correlation between CAMBUS and regulatory elements

We evaluated the functional enrichment of CAMBUS+ OCRs in known regulatory elements, including ENCODE cCREs, FANTOM5 enhancers, and super-enhancers (see URLs). For each CAMBUS+ OCR, we counted overlapping annotations for each known regulatory element and calculated the OR compared with all OCRs, and tested it using Fisher's exact test, as described in the previous section. For tissue-expressed FANTOM5 enhancers, we obtained count per million (CPM) data for 249,599 enhancers across 347 tissue/cell types with hg19 coordinates (see URLs) and selected enhancers expressed in hematopoietic_cell (N = 163), brain (N = 85), kidney (N = 16), and lung (N = 8) based on a threshold of CPM >= 0.1 in at least 10% of individuals in each enhancer. As described previously, the positions were mapped to the hg38/GRCh38 coordinate using liftOver. For tissue-

expressed super-enhancers, we obtained Blood (N = 115), brain (N = 203), kidney (N = 34), and lung (N = 119) enhancers with hg38 coordinates (see URLs) and merged these enhancers by tissue with `bedtools merge` function. The adjusted OR with Gnocchi, phyloP46way, phyloP100way, cactus241way, phyloPPrimates, GERP, and DR was estimated using a logistic regression model, adjusting for the average of these seven constraint metrics overlapping CAMBUS+ OCRs and control CAMBUS OCRs.

For all analyses, regions were considered to be overlapped if they had more than 50% overlap with each other.

Heritability enrichment analysis

To estimate heritability enrichment in each OCR, we conducted stratified-linkage disequilibrium score regression (S-LDSC) with LDSC (ver.1.0.1) using the baseline model (version 1.2; 53 categories). The p-value of the coefficient for CAMBUS+ OCR (or relaxed CAMBUS+ OCR) was obtained by adjusting for the 53 categories in the baseline model and all OCRs in each cell type. To avoid false positives (Tashman, Cui et al. 2021), we used the relaxed CAMBUS+ OCRs (the median proportion of single nucleotide polymorphism (SNPs) across the 29 cell types was 0.5%) for primary analyses instead of the CAMBUS+ OCRs (the median was 0.07%). Multiple testing correction across all trait-cell type pairs was performed to estimate FDR using the Storey method (Storey 2002). The summary statistics were obtained from Alkes Price Lab (See URLs and our previous study (Oguchi, Suzuki et al. 2024)), and five GWAS results in EAS ancestry: systemic lupus erythematosus (Yin, Kim et al. 2021), rheumatoid arthritis (Ishigaki, Sakaue et al. 2022), Systemic sclerosis (Ishikawa, Tanaka et al. 2024), Atopic dermatitis (Tanaka, Koido et al. 2021), and Takayasu arteritis (Terao, Yoshifuji et al. 2018). For this analysis, CAMBUS

OCRs' positions were mapped to the hg19 coordinate using the UCSC liftOver tool with hg38ToHg19.over.chain file to be consistent with the summary statistics. After mapping to hg19, regions mapped to different chromosomes were excluded from the analysis. For the other constraint metrics, we extracted the strong constraints region: Gnocchi Z > 4, phyloP46way score > 0.655, phyloP100way score > 9, phyloPPrimates score > 0.999, cactus241way score > 8.8, GERP RS > 5, DR rank <= 0.005. These thresholds were adjusted so that the proportion of SNPs matched that of Gnocchi (1.18%) because using the recommended thresholds (FDR 5% for phyloP (Sullivan, Meadows et al. 2023), 0.961 for phyloPPrimates (Sullivan, Meadows et al. 2023), 2 for GERP (a RS score threshold of 2 provides high sensitivity while still strongly enriching for truly constrained sites, as described in the UCSC Genome Browser documentation at 2025/7/22) (Fu, Liu et al. 2014), 0.01 for DR (Halldorsson, Eggertsson et al. 2022), cactus241way with score > 2.27 (FDR 5%)) resulted in a proportion of SNPs that was too large, leading to lower resolution to assess heritability enrichment. For example, the recommended parameter for cactus241way captured 66% of the proportion of SNPs and yielded only modest SNP heritability enrichment (e.g., 1.37 for SLE), while the threshold that we used (8.8) substantially improved the enrichment (7.32 for SLE).

PRS analysis

The PRS was constructed using PLINK v1.9, following a systematic approach with p-value thresholds using SLE GWAS in EAS ancestry (1,178 cases and 136,442 controls, in total, 137,620 individuals) (Yin, Kim et al. 2021). To select relevant variants for the PRS, we selected SNPs that met 20 types of p-value thresholds: 1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.05, 0.01, 0.005, 5.0×10^{-4}, 5.0×10^{-5}, 1.0×10^{-5}, 5.0×10^{-6}, 1.0×10^{-7}, 5.0×10^{-7} and 5.0×10^{-8}. After selecting SNPs based on these thresholds, we applied

LD clumping to remove SNPs in high linkage disequilibrium (LD) using nine types of r^2 thresholds (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9) and a 250kb physical distance. This process yielded 180 PRS models. To determine the optimal PRS model, we applied a 5-fold cross-validation strategy. The dataset was divided into five folds, with four folds used for training and one for testing in each iteration. We computed PRSs for all 180 models using variants from SLE GWAS in EAS ancestry for each training set. We calculated the AUC with pROC (v1.18.4) to select the combination of p-value and r^2 parameters that yielded the best AUC for each training set. We then took the two variant sets; (a) variants filtered with the best parameters, and (b) all variants overlapping with DN B CAMBUS+ OCRs as prioritized variants. To mitigate overfitting due to LD structures, we removed the variants with LD>0.7 with variants in set (b). We then calculated the PRS for both sets (a) and (b) for each of the five test sets, and estimated the effect size of PRS from set (b), controlling for the PRS from set (a) using a logistic regression model. Finally, we performed a fixed-effects meta-analysis to evaluate the improvement in predictive power with the prioritized variants.

Fine-mapping enrichment analysis

We downloaded the UKB fine-mapped variants of 94 complex diseases and traits from the Finucane lab (see URLs) and obtained BBJ fine-mapped variants of 62 quantitative traits from a recent study (Koyama, Liu et al. 2024). Regarding the UKB fine-mapped variants, we used PIP from the SuSiE. The positions of those datasets were mapped to the hg38/GRCh38 coordinate using liftOver as described previously. In total, we obtained 830,654 trait-variant pairs for UKB and 544,745 trait-variant pairs for BBJ. The number of overlapped variants for each trait in 29 cell types of CAMBUS+ OCRs was counted, and the OR was estimated using all OCRs in each 29-cell type as the background, and the P-value was calculated from the logistic regression model. After Bonferroni correction (P=0.05

divided by the number of tested phenotypes), the median of OR is calculated from the significant enrichment in UKB and BBJ (Extended Data Fig. 4A, B; Supplementary Table 20).

We categorized each trait into hematopoietic and non-hematopoietic. Furthermore, hematopoietic traits are categorized into leukocyte and non-leukocyte traits. The classification of traits and the number of putative causal variants in each category are described in Supplementary Table 21.

Then, we counted the overlap of variants for each leukocyte and non-leukocyte and non-hematopoietic for each 29 cell-type CAMBUS+ OCRs. We calculated the OR using the method described above.

To perform enrichment analysis of rare variants, we first created merged OCRs by combining CAMBUS+ OCRs from 29 different cell types (merged CAMBUS+ OCRs). The number of overlapped variants with MAF < 0.01 and PIP > 0.1 in merged CAMBUS+ OCRs was counted for all traits and only hematopoietic traits, and the OR was calculated using the variants with MAF < 0.01 and PIP <= 0.1 as the background and calculated one-sided Fisher's exact P-value.

Catalog cell-type-specific putative causal variants

We chose variants as putative causal variants from each of the credible sets overlapping only one variant with CAMBUS+ OCRs. Then, we filtered the variants that overlapped with only one cell type of CAMBUS+ OCRs. In total, we obtained 186 and 62 cell-type-specific putative causal variants. Furthermore, we chose the variants that overlapped only one subgroup (one of the CD4, CD8, B, Mono, DC, NK, Neutrophil; Supplementary Table 1).

Finally, we obtained 1,210 and 323 for trait-related cell-type-specific putative causal variants from both EUR and EAS ancestries, respectively.

Statistical analysis

For p-value calculation with R, in case of insufficient precision with standard floating-point arithmetic, the Rmpfr package (ver 0.9.2) was used to handle high-precision floating-point arithmetic to calculate p-value with z-score. For the meta-analysis, the pooled OR was calculated using a random effect model with metafor R package (version 2.0-0).

Data availability

The CAMBUS will be publicly available online soon after acceptance.

Code availability

The CAMBUS and its calculation scripts will be publicly available online soon after acceptance.

URLs

Gnocchi is at https://gnomad.broadinstitute.org/downloads#v3-genomic-constraint. Enrichment for Gnocchi with the known cCREs is at gs://gnomad-nc-constraint-v31-paper/fig_tables/constraint_z_genome_1kb.annot.txt. ENCODE cCREs V3 is at https://screen.encodeproject.org/. FANTOM5 enhancer is at https://zenodo.org/record/556775/. Super enhancer is at https://bio.liclab.net/sedb/download.php. phylop46way is at https://hgdownload.cse.ucsc.edu/goldenPath/hg19/phyloP46way/primates. phyloP100way is at

https://hgdownload.cse.ucsc.edu/goldenpath/hg38/phyloP100way/hg38.phyloP100way.bw phyloPPrimates is at https://cgl.gi.ucsc.edu/data/cactus/zoonomia-2021-track-hub/hg38/phyloPPrimates.bigWig. cactus241way is at https://hgdownload.soe.ucsc.edu/goldenPath/hg38/cactus241way/cactus241way.phyloP.bw. GERP is at ftp.ensembl.org/pub/current_compara/conservation_scores/91_mammals.gerp_conservation_score/gerp_conservation_scores.homo_sapiens.GRCh38.bw. DR is at https://static-content.springer.com/esm/art%3A10.1038%2Fs41586-022-04965-x/MediaObjects/41586_2022_4965_MOESM3_ESM.gz.GWAS summary statistics are at https://alkesgroup.broadinstitute.org/sumstats_formatted, and https://alkesgroup.broadinstitute.org/UKBB/. ABC predictions are at https://mitra.stanford.edu/engreitz/oak/public/Nasser2021/AllPredictions.AvgHiC.ABC0.015.minus150.ForABCPaperV3.txt.gz. UKB fine-mapped variants are at https://www.finucanelab.org/data. FANTOM5 tissue-expressed data is at https://zenodo.org/records/5348471.

Acknowledgment

We would like to thank all the participants in the study. We deeply thank all the members of the Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences for the technical support. We acknowledge the use of Biorender for aspects of the figures. This work was supported by Japan Agency for Medical Research and Development (AMED) grants 21ek0109555, 21tm0424220, 21ck0106642, 23ek0410114, and 23tm0424225. Japan Society for the Promotion of Science (JSPS) KAKENHI grant JP20H00462 and JP23K14164, Takeda Hosho Grants for Research in Medicine.

Author contributions

K.T., M.K., and C.T. conceived the study design. A.S. performed ATAC-sequencing and DNA genotyping. K.T. and N.T. performed whole genome imputation. K.T. analyzed this study with the help of S.Y., N.T., Y.I., X.L., S.K., and M.K. M.K. and C.T. supervised this study. S.Y., N.T., S.K., K.I., Y.M., Immune transcript/enhancer consortium, and K.Y. provided helpful comments and discussion. K.T., M.K., A.S., and C.T. wrote the manuscript.

Ethics declarations

Competing interests

The authors declare no competing financial interests.

Figure legends

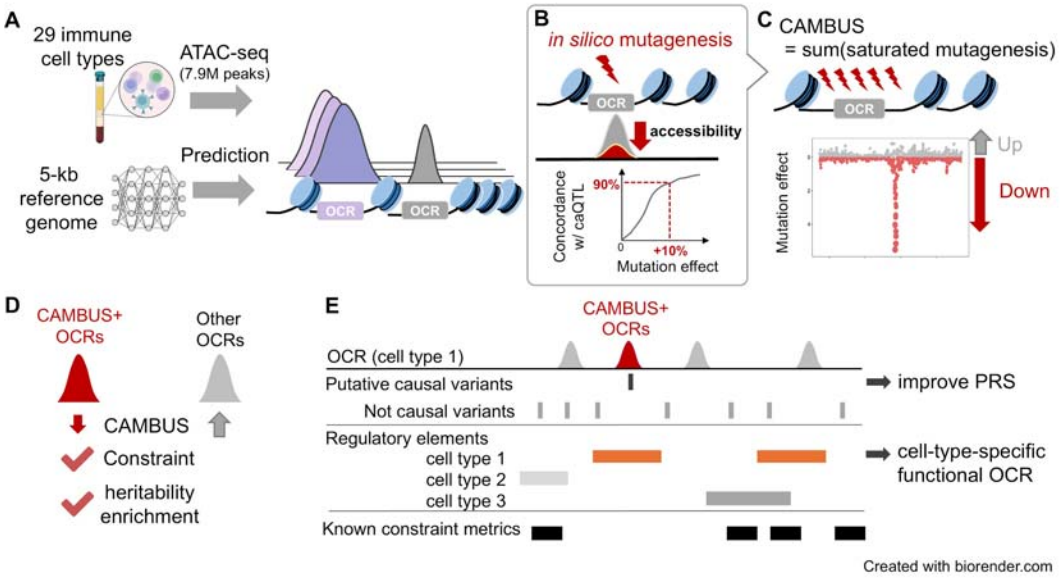


Fig. 1 | Study design. Overview of the study design. **A,** Dataset comprising 29 cell types was used to validate our approach, and a machine learning model was developed to predict OCRs. **B,C,** Identification of putative constraint OCRs (CAMBUS+). **D,** Characteristics of CAMBUS+ OCRs based on CAMBUS. **E,** Characterization of CAMBUS+ OCRs and prioritization for putative causal variants.

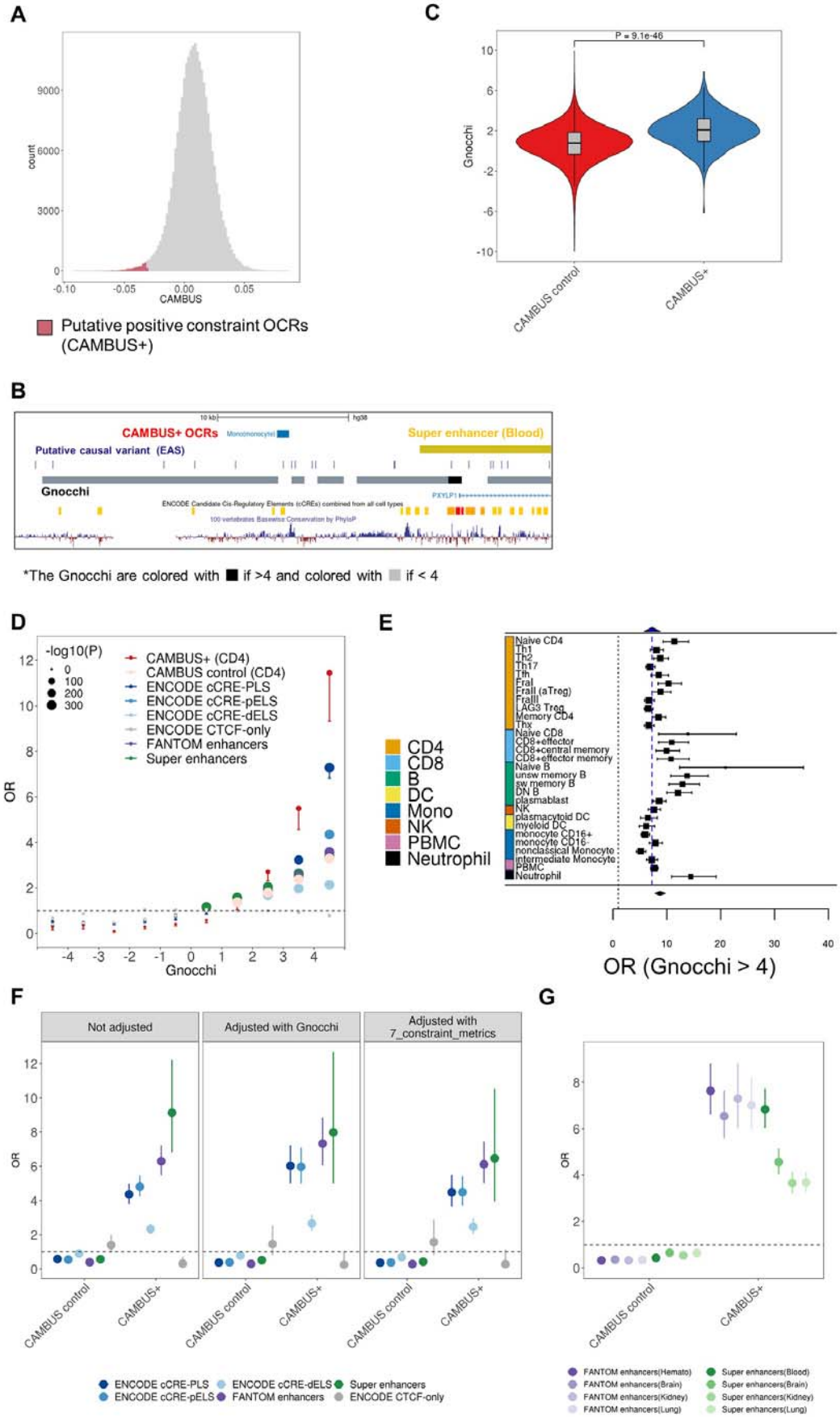

A
count
CAMBUS
Putative positive constraint OCRs (CAMBUS+)
C
P = 9.1e-46
Gnocchi
CAMBUS control
CAMBUS+
B
CAMBUS+ OCRs
Super enhancer (Blood)
Putative causal variant (EAS)
Gnocchi
*The Gnocchi are colored with ■ if >4 and colored with ■ if < 4
D
-log10(P)
CAMBUS+ (CD4)
CAMBUS control (CD4)
ENCODE cCRE-PLS
ENCODE cCRE-pELS
ENCODE cCRE-dELS
ENCODE CTCF-only
FANTOM enhancers
Super enhancers
OR
Gnocchi
E
CD4
CD8
B
DC
Mono
NK
PBMC
Neutrophil
OR (Gnocchi > 4)
F
Not adjusted
Adjusted with Gnocchi
Adjusted with 7_constraint_metrics
OR
CAMBUS control
CAMBUS+
ENCODE cCRE-PLS
ENCODE cCRE-pELS
ENCODE cCRE-dELS
FANTOM enhancers
Super enhancers
ENCODE CTCF-only
G
OR
CAMBUS control
CAMBUS+
FANTOM enhancers(Hemato)
FANTOM enhancers(Brain)
FANTOM enhancers(Kidney)
FANTOM enhancers(Lung)
Super enhancers(Blood)
Super enhancers(Brain)
Super enhancers(Kidney)
Super enhancers(Lung)

Fig. 2 | CAMBUS+ OCRs can mark constraint regions and active enhancers. A, Distribution of CAMBUS. The red part represents putative constraint OCRs. **B,** Genome browser view at representative constraint region around CAMBUS+ in monocytes (chr3:141,217,899-141,218,844). The black box represents Gnocchi with Z > 4, and the grey box represents Gnocchi with Z < 4. **C,** Distribution of non-coding Gnocchi in CAMBUS+ OCRs and CAMBUS control in NaiveCD4. The one-sided Wilcoxon-tests were performed between CAMBUS+ OCRs and CAMBUS+ control. **D,E,** Enrichment (odds ratio; OR) of CAMBUS+ in non-coding Gnocchi (n = 1,843,599), comparing with the known regulatory elements (using the previous study's results, see URLs) in non-coding Gnocchi (n = 1,843,599). We calculated the OR of CAMBUS+ (n = 1,266) and CAMBUS control (n = 251,807) for NaiveCD4 using various Gnocchi's thresholds in (**D**) and OR for each CAMBUS+ OCR in 29 cell types and their pooled OR through random effects meta-analysis using Gnocchi > 4 in (**E**). The blue dashed line represents the OR of ENCODE cCRE-PLS (**E**). **F**, Enrichment of CAMBUS+ OCRs for NaiveCD4 (n = 1,266) in the known cCREs (ENCODE cCRE-PLS; n = 40,891, ENCODE cCRE-pELS; n = 172,027, ENCODE cCRE-dELS; n = 789,200, ENCODE CTCF-only; n = 35,839, FANTOM5 enhancers; n = 63,280, Super enhancers; n = 16,472), adjusted for Gnocchi, phyloP46way, phyloP100way, cactus241way, phyloPPrimates, GERP and DR. **G,** Enrichment of CAMBUS+ for NaiveCD4 (n = 1,266) in the tissue expressed FANTOM5 enhancers (hematopoietic cells; n = 64,225, Brain; n = 62,978, Kidney; n = 35,177, Lung; n = 26,630), and Super enhancers (Blood; n = 8,055, Brain; n = 9,372, Kidney; n = 4,712, Lung; n = 9,878). The dashed line represents no enrichment (OR=1.0; **D, E, F, G**). CAMBUS control in NaiveCD4 (n = 251,807) is used as background in (**F, G**). The error bars represent 95% confidence intervals, with only lower bounds shown in **D** (See Supplementary Table 4 for higher bounds).

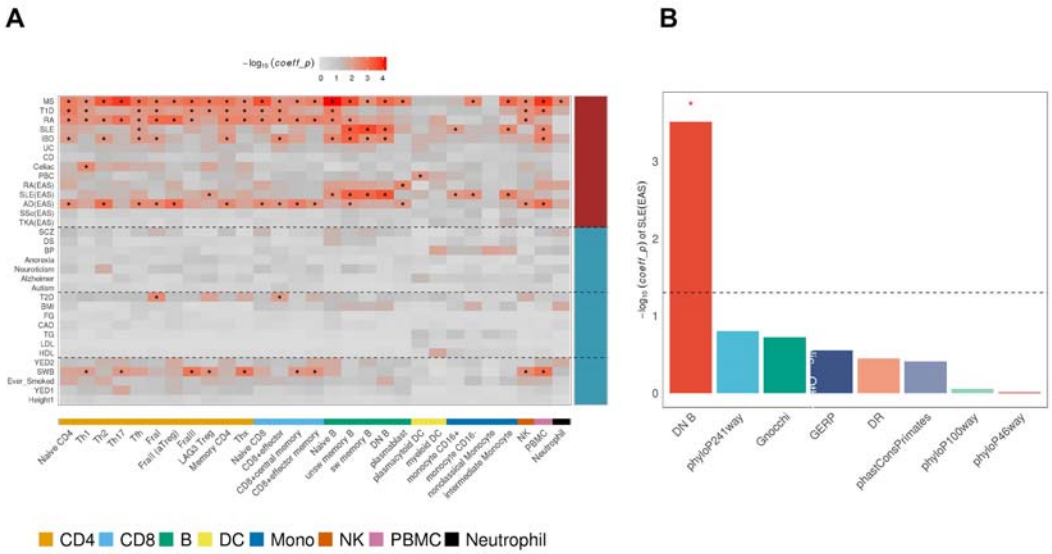
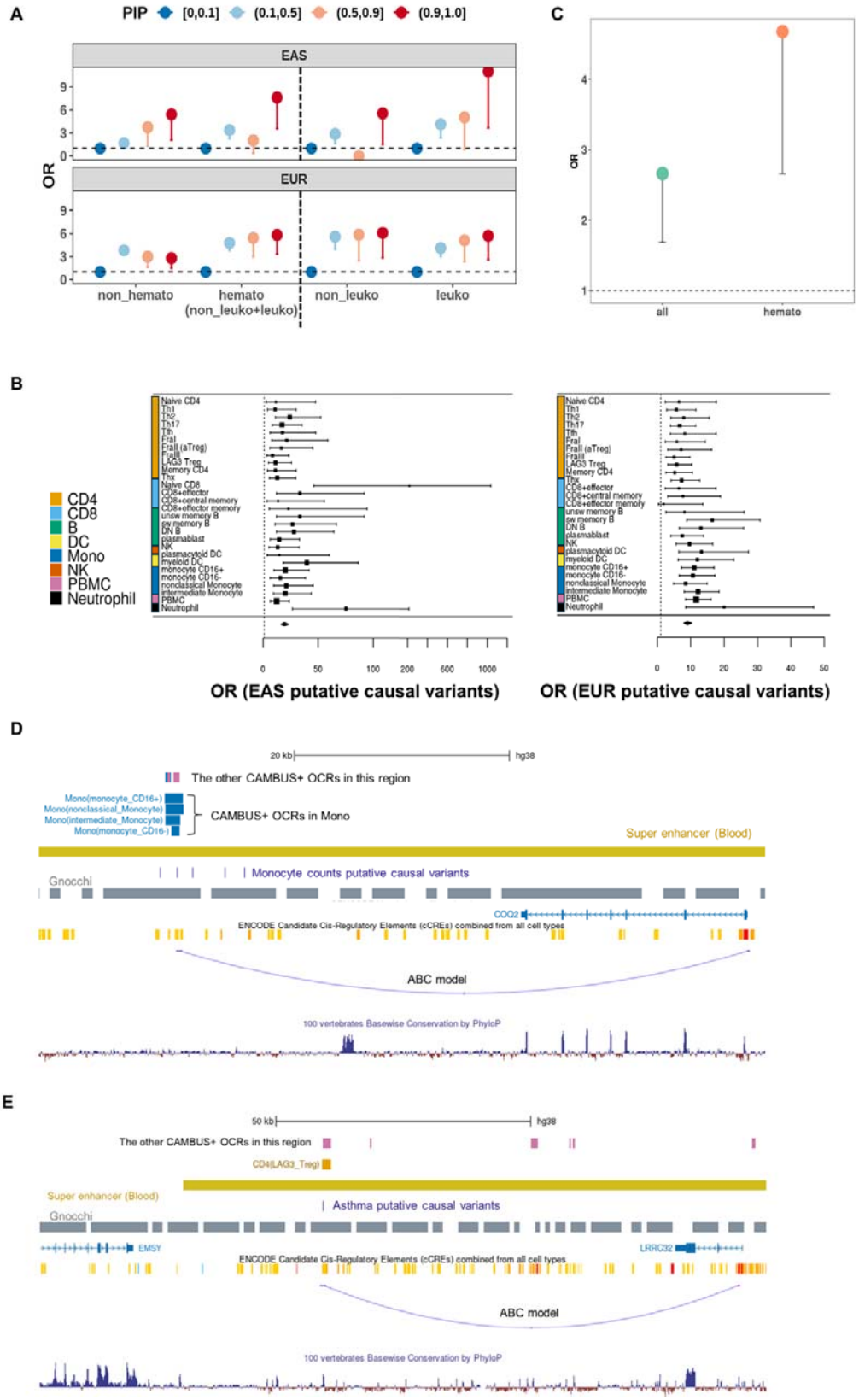


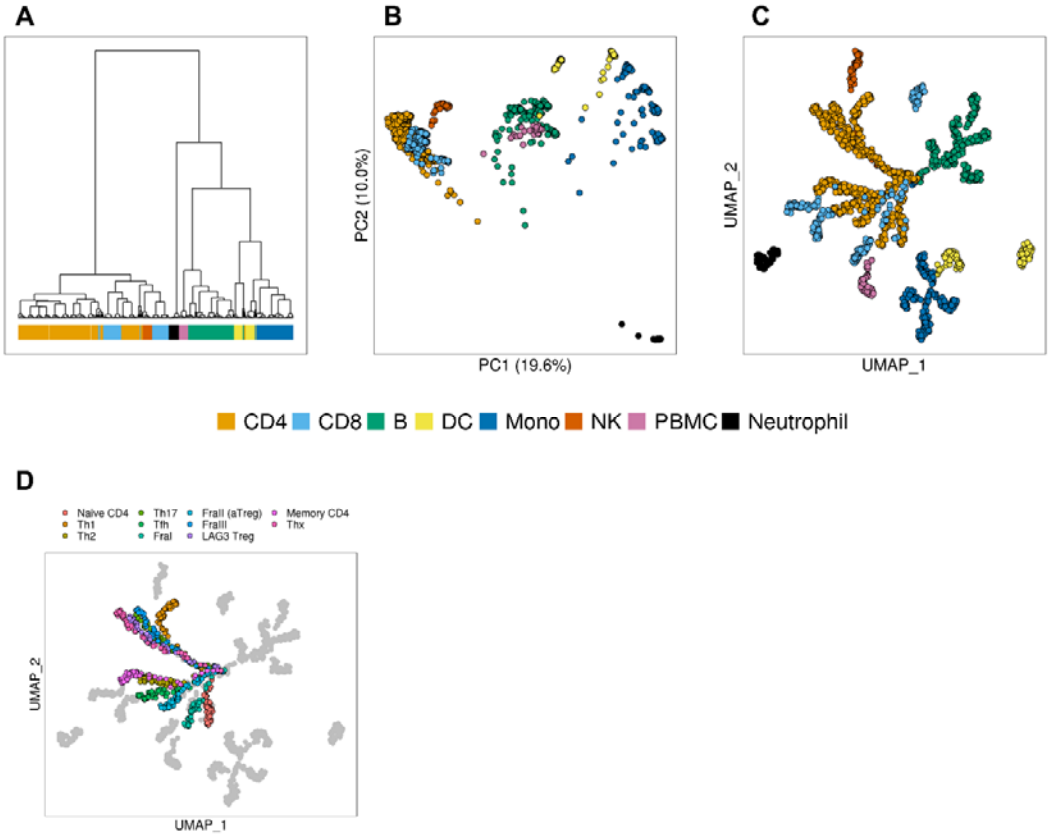
Fig. 3 | Heritability enrichment for immune-mediated diseases. A, The heritability enrichment of various GWAS traits (in total, 33 GWAS for EUR and EAS ancestries; Supplementary Table 13) in CAMBUS+ OCRs for each of 29 cell types. The heatmap colors indicate -log10-scale p-value of the coefficient for CAMBUS+ annotation in S-LDSC (coefficient P). **B,** The -log10-scale coefficient P of CAMBUS+ OCRs in double-negative B-cells (DN B) and seven known constraint metrics for SLE GWAS in EAS ancestry. '*' indicates significance (qvalue < 0.05) after FDR correction (**A, B**).


A
PIP [0,0.1] (0.1,0.5] (0.5,0.9] (0.9,1.0]
EAS
EUR
OR
non_hemato
hemato (non_leuko+leuko)
non_leuko
leuko
C
OR
all
hemato
B
CD4
CD8
B
DC
Mono
NK
PBMC
Neutrophil
OR (EAS putative causal variants)
OR (EUR putative causal variants)
D
20 kb
hg38
The other CAMBUS+ OCRs in this region
Mono(monocyte_CD16+)
Mono(nonclassical_Monocyte)
Mono(intermediate_Monocyte)
Mono(monocyte_CD16-)
CAMBUS+ OCRs in Mono
Super enhancer (Blood)
Monocyte counts putative causal variants
Gnocchi
COQ2
ENCODE Candidate Cis-Regulatory Elements (cCREs) combined from all cell types
ABC model
100 vertebrates Basewise Conservation by PhyloP
E
50 kb
hg38
The other CAMBUS+ OCRs in this region
CD4(LAG3_Treg)
Super enhancer (Blood)
Asthma putative causal variants
Gnocchi
EMSY
LRRC32
ENCODE Candidate Cis-Regulatory Elements (cCREs) combined from all cell types
ABC model
100 vertebrates Basewise Conservation by PhyloP

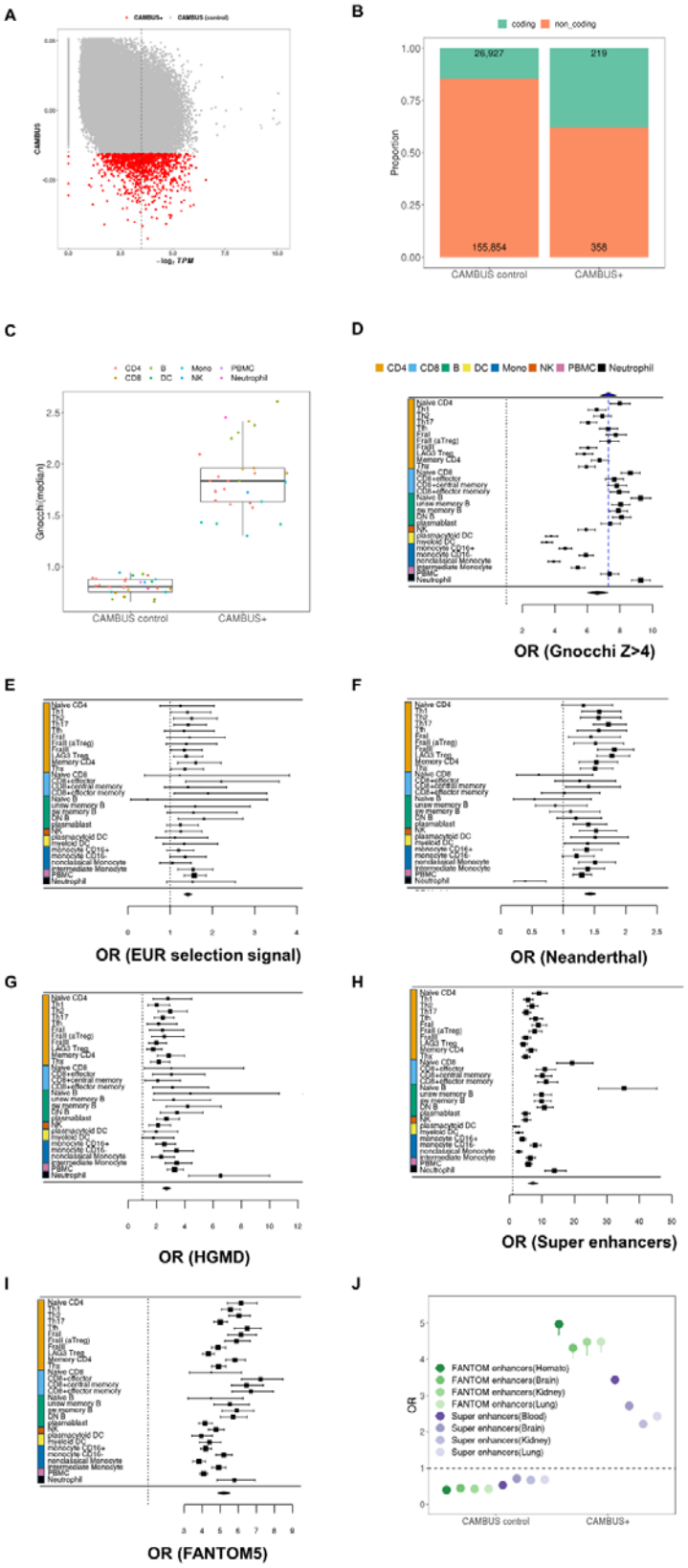
750

Fig. 4 | Prioritization of putative causal variants using CAMBUS+ OCRs. A, Enrichment of putative causal variants in leukocyte (leuko), non-leukocytes (non_leuko), hemato (leuko + non_leuko), and non-hematopoietic (non_hemato) traits in CAMBUS+ OCRs for Th1. Putative causal variants were obtained from statistical fine-mapping analysis of many complex traits' GWAS results for EAS and EUR populations (Kanai, Ulirsch et al. 2021, Koyama, Liu et al. 2024). **B,** Forest plot showing the enrichment of CAMBUS+ OCRs across 29 cell types and the pooled OR in highly confident putative causal variants (PIP > 0.9). Variants with PIP <= 0.1 were used as the background for OR calculation. **C,** Enrichment of CAMBUS+ OCRs for any of 29 cell types in putative causal rare variants (maf < 0.01 in EAS), evaluated by comparing the proportion of CAMBUS+ OCRs overlapping with putative causal rare variants with PIP > 0.1. The left dot represents the OR calculated using all the putative causal variants, while the right dot represents the OR based only on the putative causal variants for hematopoietic phenotypes. The dashed line represents no enrichment (OR=1.0). **D,E,** Representative examples of variants associated with monocyte counts prioritized with CAMBUS+ in Mono (**D**) and asthma prioritized with CAMBUS+ in regulatory T-cells (**E**) in putative causal variants in EUR ancestry and surrounding genomic annotations. The navy boxes represent putative causal variants in this credible set. The blue dashed line represents an enhancer-gene connection to the nearest protein-coding gene predicted by the ABC model for $CD14^+$ monocytes (**D**) and HUVEC (**E**). The gene with the highest ABC score is also shown in (**D**). All the other CAMBUS+ OCRs around putative causal variants (rs13122305 (**D**), rs55646091 (**E**)) are also shown in the figures (The purple box represents PBMC).

Extended Data Figure

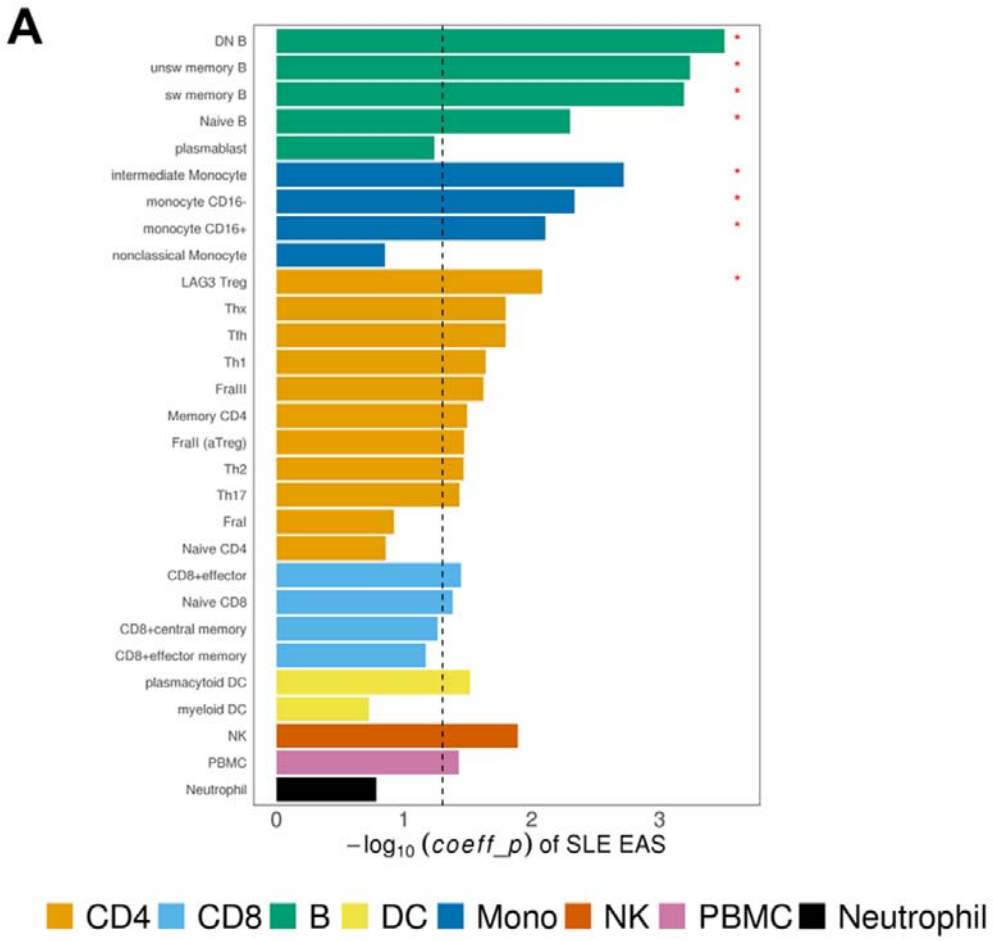


Extended Data Fig. 1 | Clustering analysis and dimension reduction using PCA and UMAP. A-C, (**A**) Hierarchical clustering, (**B**) PCA, and (**C**) UMAP of 1,386 samples. The top 5,000 most accessible OCRs were used for these analyses. The colors indicate the classification of samples into eight subgroups across 29 cell types. **D,** UMAP of 1,386 samples (**C**) colored by eleven types of CD4; other subgroups are colored with grey.

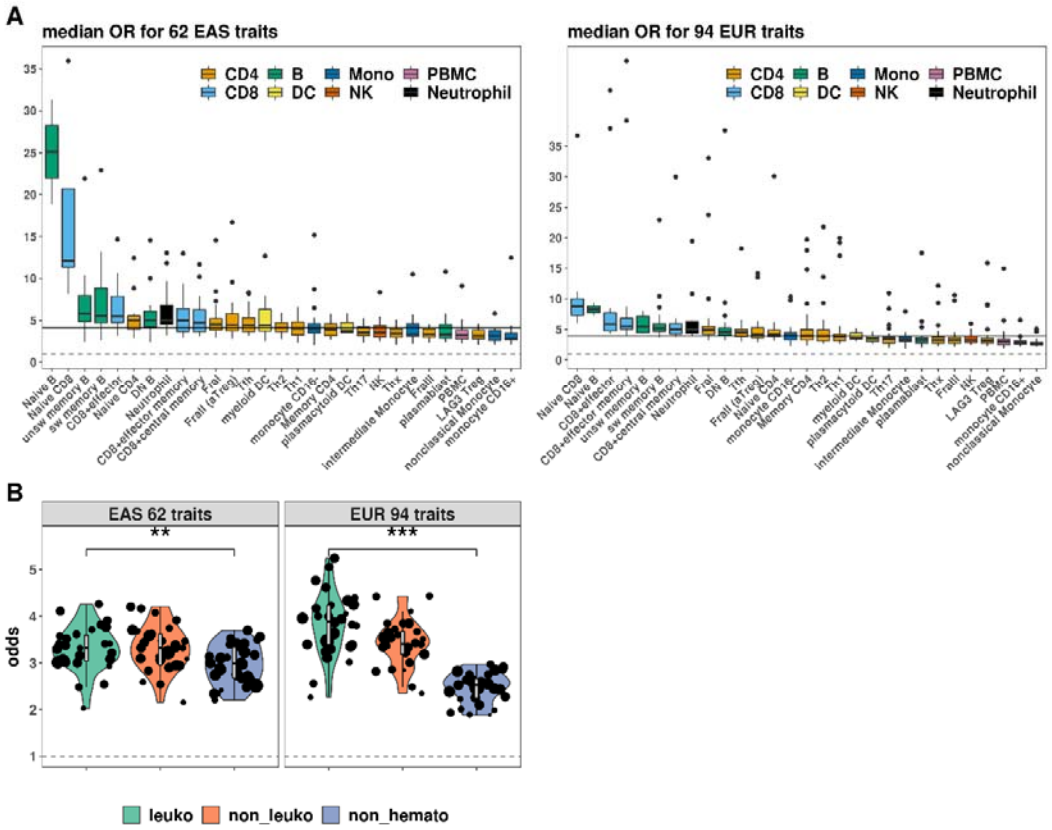

A
B
coding
non_coding
26,927
219
155,854
358
Proportion
CAMBUS control
CAMBUS+
C
D
CD4
CD8
B
DC
Mono
NK
PBMC
Neutrophil
Gnocchi (median)
CAMBUS control
CAMBUS+
OR (Gnocchi Z>4)
E
F
OR (EUR selection signal)
OR (Neanderthal)
G
H
OR (HGMD)
OR (Super enhancers)
I
J
OR (FANTOM5)
FANTOM enhancers(Hemato)
FANTOM enhancers(Brain)
FANTOM enhancers(Kidney)
FANTOM enhancers(Lung)
Super enhancers(Blood)
Super enhancers(Brain)
Super enhancers(Kidney)
Super enhancers(Lung)
OR
CAMBUS control
CAMBUS+

782

Extended Data Fig. 2 | CAMBUS+ can mark constraint regions and active enhancers. **A,** Scatter plot of median TPM and CAMBUS in Naïve CD4. The vertical dashed line represents the top 5% of the median TPM. **B,** The number of OCRs in coding and non-coding regions for each CAMBUS+ OCRs and CAMBUS control. The regions of coding and non-coding are referred to previous study (Chen, Francioli et al. 2024). **C,** The distribution of non-coding Gnocchi (median) for each CAMBUS+ OCRs and CAMBUS control in each 29 cell type. **D,** OR of relaxed CAMBUS+ OCRs across 29 cell types in Gnocchi> 4 and pooled OR estimated through random effect meta-analysis. The blue dashed line represents the OR of ENCODE promoter-like elements. **E-I,** OR and the pooled OR of the CAMBUS+ OCRs in UKB selection signals (**E**), Neanderthal genome sequences derived from samples of EAS ancestry (**F**), pathogenic regulatory variants (**G**), super-enhancers (**H**) and FANTOM5 enhancers (**I**). **J,** ORs of relaxed CAMBUS+ (NaiveCD4; n = 12,736) in cell-type-specific regulatory elements: the tissue expressed FANTOM5 enhancers (hematopoietic cells; n = 64,225, Brain; n = 62,978, Kidney; n = 35,177, Lung; n = 26,630), and super-enhancers (Blood; n = 8,055, Brain; n = 9,372, Kidney; n = 4,712, Lung; n = 9,878). The dashed line represents no enrichment (OR=1.0; **D-J**)

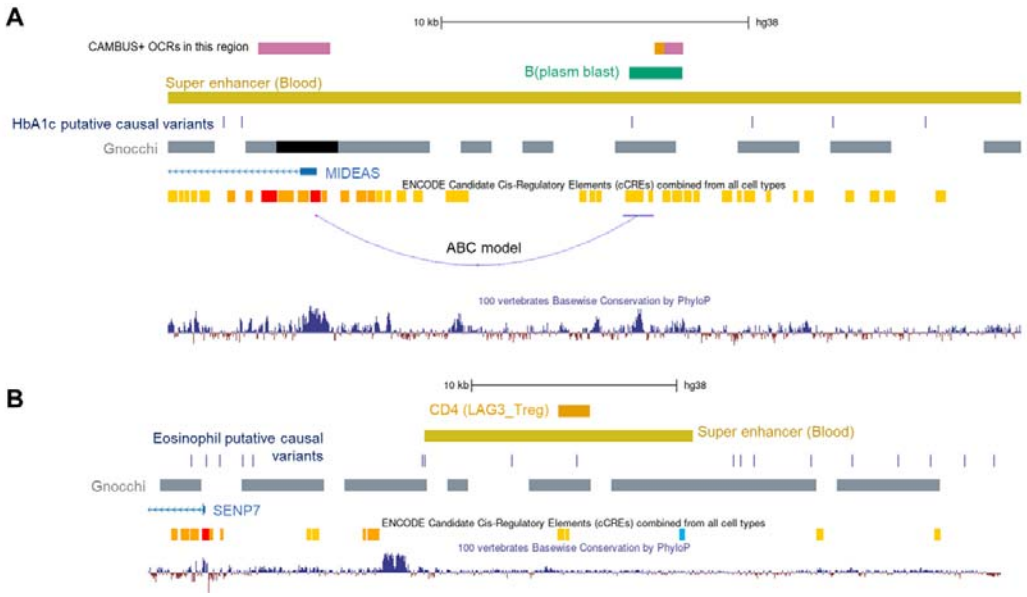


Extended Data Fig. 3 | Heritability enrichment for immune-mediated diseases. A, The -log10-scale coefficient P of CAMBUS+ OCRs in 29 cell types for SLE GWAS in EAS ancestry. The dashed line indicates a coefficient P = 0.05. '*' indicates significance (qvalue < 0.05) after FDR correction.

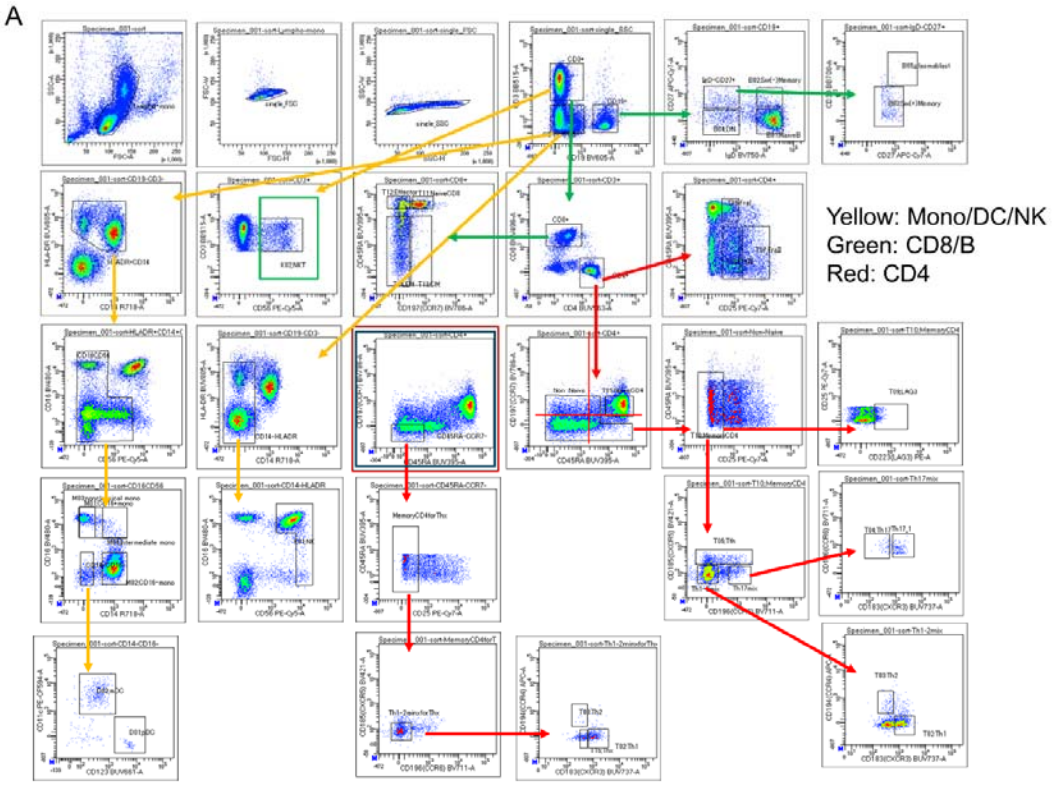


Extended Data Fig. 4 | Prioritization of putative causal variants using CAMBUS+ OCRs. A, Median OR of all variants in 95% credible sets in CAMBUS+ OCRs across 29 cell types. We used putative causal variants from statistical fine-mapping studies in EUR and EAS ancestries (Kanai, Ulirsch et al. 2021, Koyama, Liu et al. 2024). The solid line represents the median OR across 29 cell types. **B,** Enrichment of the putative causal variants in leukocyte (leuko), non-leukocytes (non_leuko), and non-hematopoietic (non_hemato) traits in CAMBUS+ OCRs across 29 cell types. Only cell types with $P < 0.05$ are shown (27 cell types in BBJ leuko; 28 in UKB non_hemato; 29 in BBJ non_leuko, BBJ non_hemato, UKB leuko, and UKB non_leuko), and dots indicate -log10-scale p-value. t-

tests were performed between leuko and non_hemato traits, with significance levels denoted as P <0.05 ('*') and P < 0.001 ('***').



Extended Data Fig. 5 | Prioritization of putative causal variants using CAMBUS+ OCRs. A,B, representative example of variants associated with Hemoglobin A1c level prioritized with CAMBUS+ in plasma blast in putative causal variants in EUR ancestry (**A**) and eosinophils prioritized with CAMBUS+ in regulatory T-cell (**B**) and their surrounding genomic annotations. The navy boxes represent putative causal variants in the credible set. The blue dashed line represents an enhancer-gene connection to the nearest protein-coding gene predicted by the ABC model in B-cells (**A**). All the other CAMBUS+ OCRs around putative causal variants (rs8009224 (**A**)) are also shown in the figures (The purple box represents PBMC).



Extended Data Fig. 6 | Sorting strategy for immune cell subsets. A, Representative FACS plots describing gating strategies for 27 cell types. See Supplementary Table 22 for flow cytometry antibodies used in this study.

References

Andersson, R., C. Gebhard, I. Miguel-Escalada, I. Hoof, J. Bornholdt, M. Boyd, Y. Chen, X. Zhao, C. Schmidl, T. Suzuki, E. Ntini, E. Arner, E. Valen, K. Li, L. Schwarzfischer, D. Glatz, J. Raithel, B. Lilje, N. Rapin, F. O. Bagger, M. Jorgensen, P. R. Andersen, N. Bertin, O. Rackham, A. M. Burroughs, J. K. Baillie, Y. Ishizu, Y. Shimizu, E. Furuhata, S. Maeda, Y. Negishi, C. J. Mungall, T. F. Meehan, T. Lassmann, M. Itoh, H. Kawaji, N. Kondo, J. Kawai, A. Lennartsson, C. O. Daub, P. Heutink, D. A. Hume, T. H. Jensen, H. Suzuki, Y. Hayashizaki, F. Muller, A. R. R. Forrest, P. Carninci, M. Rehli and A. Sandelin (2014). "An atlas of active enhancers across human cell types and tissues." Nature **507**(7493): 455-461.

Armanfar, M., A. Jafari, G. R. Dehghan and L. Abdizadeh (2015). "Effect of coenzyme Q10 supplementation on exercise-induced response of inflammatory indicators and blood lactate in male runners." Med J Islam Repub Iran **29**: 202.

Barton, A. R., M. A. Sherman, R. E. Mukamel and P. R. Loh (2021). "Whole-exome imputation within UK Biobank powers rare coding variant association and fine-mapping analyses." Nat Genet **53**(8): 1260-1269.

Benaglio, P., J. Newsome, J. Y. Han, J. Chiou, A. Aylward, S. Corban, M. Miller, M. L. Okino, J. Kaur, S. Preissl, D. U. Gorkin and K. J. Gaulton (2023). "Mapping genetic effects on cell type-specific chromatin accessibility and annotating complex immune trait variants using single nucleus ATAC-seq in peripheral blood." PLoS Genet **19**(6): e1010759.

Boyle, A. P., S. Davis, H. P. Shulha, P. Meltzer, E. H. Margulies, Z. Weng, T. S. Furey and G. E. Crawford (2008). "High-resolution mapping and characterization of open chromatin across the genome." Cell **132**(2): 311-322.

Brixi, G., M. G. Durrant, J. Ku, M. Poli, G. Brockman, D. Chang, G. A. Gonzalez, S. H. King, D. B. Li, A. T. Merchant, M. Naghipourfar, E. Nguyen, C. Ricci-Tam, D. W. Romero, G. Sun, A. Taghibakshi, A. Vorontsov, B. Yang, M. Deng, L. Gorton, N. Nguyen, N. K. Wang, E. Adams, S. A. Baccus, S. Dillmann, S. Ermon, D. Guo, R. Ilango, K. Janik, A. X. Lu, R. Mehta, M. R. K. Mofrad, M. Y. Ng, J. Pannu, C. Ré, J. C. Schmok, J. S. John, J. Sullivan, K. Zhu, G. Zynda, D. Balsam, P. Collison, A. B. Costa, T. Hernandez-Boussard, E. Ho, M.-Y. Liu, T. McGrath, K. Powell, D. P. Burke, H. Goodarzi, P. D. Hsu and B. L. Hie (2025). "Genome modeling and design across all domains of life with Evo 2." bioRxiv: 2025.2002.2018.638918.

Browning, S. R., B. L. Browning, Y. Zhou, S. Tucci and J. M. Akey (2018). "Analysis of Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture." Cell **173**(1): 53-61 e59.

Buenrostro, J. D., P. G. Giresi, L. C. Zaba, H. Y. Chang and W. J. Greenleaf (2013). "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position." Nat Methods **10**(12): 1213-1218.

Chen, S., L. C. Francioli, J. K. Goodrich, R. L. Collins, M. Kanai, Q. Wang, J. Alfoldi, N. A. Watts, C. Vittal, L. D. Gauthier, T. Poterba, M. W. Wilson, Y. Tarasova, W. Phu, R. Grant, M. T. Yohannes, Z. Koenig, Y. Farjoun, E. Banks, S. Donnelly, S. Gabriel, N. Gupta, S. Ferriera, C. Tolonen, S. Novod, L. Bergelson, D. Roazen, V. Ruano-Rubio, M. Covarrubias, C. Llanwarne, N. Petrillo, G. Wade, T. Jeandet, R. Munshi, K. Tibbetts, C. Genome Aggregation Database, A. O'Donnell-Luria, M. Solomonson, C. Seed, A. R. Martin, M. E. Talkowski, H. L. Rehm, M. J. Daly, G. Tiao, B. M. Neale, D. G. MacArthur and K. J. Karczewski (2024). "A genomic mutational constraint map using variation in 76,156 human genomes." Nature **625**(7993): 92-100.

Chung, M. K. Y., L. Gong, D. L. Kwong, V. H. Lee, A. W. Lee, X. Y. Guan, N. W. Kam and W. Dai (2023). "Functions of double-negative B cells in autoimmune diseases, infections, and cancers." EMBO Mol Med **15**(9): e17341.

Claussnitzer, M., J. H. Cho, R. Collins, N. J. Cox, E. T. Dermitzakis, M. E. Hurles, S. Kathiresan, E. E. Kenny, C. M. Lindgren, D. G. MacArthur, K. N. North, S. E. Plon, H. L. Rehm, N. Risch, C. N. Rotimi, J. Shendure, N. Soranzo and M. I. McCarthy (2020). "A brief history of human disease genetics." Nature **577**(7789): 179-189.

Consortium, E. P., J. E. Moore, M. J. Purcaro, H. E. Pratt, C. B. Epstein, N. Shoresh, J. Adrian, T. Kawli, C. A. Davis, A. Dobin, R. Kaul, J. Halow, E. L. Van Nostrand, P. Freese, D. U. Gorkin, Y. Shen, Y. He, M. Mackiewicz, F. Pauli-Behn, B. A. Williams, A. Mortazavi, C. A. Keller, X. O. Zhang, S. I. Elhajjajy, J. Huey, D. E. Dickel, V. Snetkova, X. Wei, X. Wang, J. C. Rivera-Mulia, J. Rozowsky, J. Zhang, S. B. Chhetri, J. Zhang, A. Victorsen, K. P. White, A. Visel, G. W. Yeo, C. B. Burge, E. Lecuyer, D. M. Gilbert, J. Dekker, J. Rinn, E. M. Mendenhall, J. R. Ecker, M. Kellis, R. J. Klein, W. S. Noble, A. Kundaje, R. Guigo, P. J. Farnham, J. M. Cherry, R. M. Myers, B. Ren, B. R. Graveley, M. B. Gerstein, L. A. Pennacchio, M. P. Snyder, B. E. Bernstein, B. Wold, R. C. Hardison, T. R. Gingeras, J. A. Stamatoyannopoulos and Z. Weng (2020). "Expanded encyclopaedias of DNA elements in the human and mouse genomes." Nature **583**(7818): 699-710.

Cooper, G. M., E. A. Stone, G. Asimenos, N. C. S. Program, E. D. Green, S. Batzoglou and A. Sidow (2005). "Distribution and intensity of constraint in mammalian genomic sequence." Genome Res **15**(7): 901-913.

Corces, M. R., A. E. Trevino, E. G. Hamilton, P. G. Greenside, N. A. Sinnott-Armstrong, S. Vesuna, A. T. Satpathy, A. J. Rubin, K. S. Montine, B. Wu, A. Kathiria, S. W. Cho, M. R. Mumbach, A. C. Carter, M. Kasowski, L. A. Orloff, V. I. Risca, A. Kundaje, P. A. Khavari, T. J. Montine, W. J. Greenleaf and H. Y. Chang (2017). "An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues." Nat Methods **14**(10): 959-962.

Eraslan, G., Z. Avsec, J. Gagneur and F. J. Theis (2019). "Deep learning: new computational modelling techniques for genomics." Nat Rev Genet **20**(7): 389-403.

Fillatreau, S., B. Manfroi and T. Dorner (2021). "Toll-like receptor signalling in B cells during systemic lupus erythematosus." Nat Rev Rheumatol **17**(2): 98-108.

Finucane, H. K., B. Bulik-Sullivan, A. Gusev, G. Trynka, Y. Reshef, P. R. Loh, V. Anttila, H. Xu, C. Zang, K. Farh, S. Ripke, F. R. Day, C. ReproGen, C. Schizophrenia Working Group of the Psychiatric Genomics, R. Consortium, S. Purcell, E. Stahl, S. Lindstrom, J. R. Perry, Y. Okada, S. Raychaudhuri, M. J. Daly, N. Patterson, B. M. Neale and A. L. Price (2015). "Partitioning heritability by functional annotation using genome-wide association summary statistics." Nat Genet **47**(11): 1228-1235.

Franchini, L. F. and K. S. Pollard (2017). "Human evolution: the non-coding revolution." BMC Biol **15**(1): 89.

Fu, Y., Z. Liu, S. Lou, J. Bedford, X. J. Mu, K. Y. Yip, E. Khurana and M. Gerstein (2014). "FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer." Genome Biol **15**(10): 480.

Fulco, C. P., J. Nasser, T. R. Jones, G. Munson, D. T. Bergman, V. Subramanian, S. R. Grossman, R. Anyoha, B. R. Doughty, T. A. Patwardhan, T. H. Nguyen, M. Kane, E. M. Perez, N. C. Durand, C. A. Lareau, E. K. Stamenova, E. L. Aiden, E. S. Lander and J. M. Engreitz (2019). "Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations." Nat Genet **51**(12): 1664-1669.

Gate, R. E., C. S. Cheng, A. P. Aiden, A. Siba, M. Tabaka, D. Lituiev, I. Machol, M. G. Gordon, M. Subramaniam, M. Shamim, K. L. Hougen, I. Wortman, S. C. Huang, N. C. Durand, T. Feng, P. L. De Jager, H. Y. Chang, E. L. Aiden, C. Benoist, M. A. Beer, C. J. Ye and A. Regev (2018). "Genetic determinants of co-accessible chromatin regions in activated T cells across humans." Nat Genet **50**(8): 1140-1150.

Gupta, A., K. Weinand, A. Nathan, S. Sakaue, M. J. Zhang, R. A. S. L. E. P. Accelerating Medicines Partnership, Network, L. Donlin, K. Wei, A. L. Price, T. Amariuta and S. Raychaudhuri (2023). "Dynamic regulatory elements in single-cell multimodal data implicate key immune cell states enriched for autoimmune disease heritability." Nat Genet **55**(12): 2200-2210.

Halldorsson, B. V., H. P. Eggertsson, K. H. S. Moore, H. Hauswedell, O. Eiriksson, M. O. Ulfarsson, G. Palsson, M. T. Hardarson, A. Oddsson, B. O. Jensson, S. Kristmundsdottir, B. D. Sigurpalsdottir, O. A. Stefansson, D. Beyter, G. Holley, V. Tragante, A. Gylfason, P. I. Olason, F. Zink, M. Asgeirsdottir, S. T. Sverrisson, B. Sigurdsson, S. A. Gudjonsson, G. T. Sigurdsson, G. H. Halldorsson, G. Sveinbjornsson, K. Norland, U. Styrkarsdottir, D. N. Magnusdottir, S. Snorradottir, K. Kristinsson, E. Sobech, H. Jonsson, A. J. Geirsson, I. Olafsson, P. Jonsson, O. B. Pedersen, C. Erikstrup, S. Brunak, S. R. Ostrowski, D. G. Consortium, G. Thorleifsson, F. Jonsson, P. Melsted, I. Jonsdottir, T. Rafnar, H. Holm, H. Stefansson, J. Saemundsdottir, D. F. Gudbjartsson, O. T. Magnusson, G. Masson, U. Thorsteinsdottir, A. Helgason, H. Jonsson, P. Sulem and K. Stefansson (2022). "The sequences of 150,119 genomes in the UK Biobank." Nature **607**(7920): 732-740.

Harb, H. and T. A. Chatila (2023). "Regulatory T-cells in asthma." Curr Opin Allergy Clin Immunol **23**(2): 151-157.

Hnisz, D., B. J. Abraham, T. I. Lee, A. Lau, V. Saint-André, A. A. Sigova, H. A. Hoke and R. A. Young (2013). "Super-enhancers in the control of cell identity and disease." Cell **155**(4): 934-947.

Honda, T. and K. Kabashima (2020). "Reconciling innate and acquired immunity in atopic dermatitis." J Allergy Clin Immunol **145**(4): 1136-1137.

Ishigaki, K., S. Sakaue, C. Terao, Y. Luo, K. Sonehara, K. Yamaguchi, T. Amariuta, C. L. Too, V. A. Laufer, I. C. Scott, S. Viatte, M. Takahashi, K. Ohmura, A. Murasawa, M. Hashimoto, H. Ito, M. Hammoudeh, S. A. Emadi, B. K. Masri, H. Halabi, H. Badsha, I. W. Uthman, X. Wu, L. Lin, T. Li, D. Plant, A. Barton, G. Orozco, S. M. M. Verstappen, J. Bowes, A. J. MacGregor, S. Honda, M. Koido, K. Tomizuka, Y. Kamatani, H. Tanaka, E. Tanaka, A. Suzuki, Y. Maeda, K. Yamamoto, S. Miyawaki, G. Xie, J. Zhang, C. I. Amos, E. Keystone, G. Wolbink, I. van der Horst-Bruinsma, J. Cui, K. P. Liao, R. J. Carroll, H. S. Lee, S. Y. Bang, K. A. Siminovitch, N. de Vries, L. Alfredsson, S. Rantapaa-Dahlqvist, E. W. Karlson, S. C. Bae, R. P. Kimberly, J. C. Edberg, X. Mariette, T. Huizinga, P. Dieude, M. Schneider, M. Kerick, J. C. Denny, P. BioBank Japan, K. Matsuda, K. Matsuo, T. Mimori, F. Matsuda, K. Fujio, Y. Tanaka, A. Kumanogoh, M. Traylor, C. M. Lewis, S. Eyre, H. Xu, R. Saxena, T. Arayssi, Y. Kochi, K. Ikari, M. Harigai, P. K. Gregersen, K. Yamamoto, S. Louis Bridges, Jr., L. Padyukov, J. Martin, L. Klareskog, Y. Okada and S. Raychaudhuri (2022). "Multi-ancestry genome-wide association analyses identify novel genetic mechanisms in rheumatoid arthritis." Nat Genet **54**(11): 1640-1651.

Ishikawa, Y., N. Tanaka, Y. Asano, M. Kodera, Y. Shirai, M. Akahoshi, M. Hasegawa, T. Matsushita, K. Saito, S. I. Motegi, H. Yoshifuji, A. Yoshizaki, T. Kohmoto, K. Takagi, A. Oka, M. Kanda, Y. Tanaka, Y. Ito, K. Nakano, H. Kasamatsu, A. Utsunomiya, A. Sekiguchi, H. Niiro, M. Jinnin, K. Makino, T. Makino, H. Ihn, M. Yamamoto, C. Suzuki, H. Takahashi, E. Nishida, A. Morita, T. Yamamoto, M. Fujimoto, Y. Kondo, D. Goto, T. Sumida, N. Ayuzawa, H. Yanagida, T. Horita, T. Atsumi, H. Endo, Y. Shima, A. Kumanogoh, J. Hirata, N. Otomo, H. Suetsugu, Y. Koike, K. Tomizuka, S. Yoshino, X. Liu, S. Ito, K. Hikino, A. Suzuki, Y. Momozawa, S. Ikegawa, Y. Tanaka, O. Ishikawa, K. Takehara, T. Torii, S. Sato, Y. Okada, T. Mimori, F. Matsuda, K. Matsuda, T. Amariuta, I. Imoto, K. Matsuo, M. Kuwana, Y. Kawaguchi, K. Ohmura and C. Terao (2024). "GWAS for systemic sclerosis identifies six novel susceptibility loci including one in the Fcgamma receptor region." Nat Commun **15**(1): 319.

Jarvinen, T. M., A. Hellquist, M. Zucchelli, S. Koskenmies, J. Panelius, T. Hasan, H. Julkunen, M. D'Amato and J. Kere (2012). "Replication of GWAS-identified systemic lupus erythematosus susceptibility genes affirms B-cell receptor pathway signalling and strengthens the role of IRF5 in disease susceptibility in a Northern European population." Rheumatology (Oxford) **51**(1): 87-92.

Jiang, Y., F. Qian, X. Bai, Y. Liu, Q. Wang, B. Ai, X. Han, S. Shi, J. Zhang, X. Li, Z. Tang, Q. Pan, Y. Wang, F. Wang and C. Li (2019). "SEdb: a comprehensive human super-enhancer database." Nucleic Acids Res **47**(D1): D235-D243.

Kanai, M., J. C. Ulirsch, J. Karjalainen, M. Kurki, K. J. Karczewski, E. Fauman, Q. S. Wang, H. Jacobs, F. Aguet, K. G. Ardlie, N. Kerimov, K. Alasoo, C. Benner, K. Ishigaki, S. Sakaue, S. Reilly, T. B. J. Project, FinnGen, Y. Kamatani, K. Matsuda, A. Palotie, B. M. Neale, R. Tewhey, P. C. Sabeti, Y. Okada, M. J. Daly and H. K. Finucane (2021). "Insights from complex trait fine-mapping across diverse populations." medRxiv: 2021.2009.2003.21262975.

Karczewski, K. J., L. C. Francioli, G. Tiao, B. B. Cummings, J. Alfoldi, Q. Wang, R. L. Collins, K. M. Laricchia, A. Ganna, D. P. Birnbaum, L. D. Gauthier, H. Brand, M. Solomonson, N. A. Watts, D. Rhodes, M. Singer-Berk, E. M. England, E. G. Seaby, J. A. Kosmicki, R. K. Walters, K. Tashman, Y. Farjoun, E. Banks, T. Poterba, A. Wang, C. Seed, N. Whiffin, J. X. Chong, K. E. Samocha, E. Pierce-Hoffman, Z. Zappala, A. H. O'Donnell-Luria, E. V. Minikel, B. Weisburd, M. Lek, J. S. Ware, C. Vittal, I. M. Armean, L. Bergelson, K. Cibulskis, K. M. Connolly, M. Covarrubias, S. Donnelly, S. Ferriera, S. Gabriel, J. Gentry, N. Gupta, T. Jeandet, D. Kaplan, C. Llanwarne, R. Munshi, S. Novod, N. Petrillo, D. Roazen, V. Ruano-Rubio, A. Saltzman, M. Schleicher, J. Soto, K. Tibbetts, C. Tolonen, G. Wade, M. E. Talkowski, C. Genome Aggregation Database, B. M. Neale, M. J. Daly and D. G. MacArthur (2020). "The mutational constraint spectrum quantified from variation in 141,456 humans." Nature **581**(7809): 434-443.

Klemm, S. L., Z. Shipony and W. J. Greenleaf (2019). "Chromatin accessibility and the regulatory epigenome." Nat Rev Genet **20**(4): 207-220.

Koido, M., C. C. Hon, S. Koyama, H. Kawaji, Y. Murakawa, K. Ishigaki, K. Ito, J. Sese, N. F. Parrish, Y. Kamatani, P. Carninci and C. Terao (2023). "Prediction of the cell-type-specific

transcription of non-coding RNAs from genome sequences via machine learning." Nat Biomed Eng **7**(6): 830-844.
Koyama, S., X. Liu, Y. Koike, K. Hikino, M. Koido, W. Li, K. Akaki, K. Tomizuka, S. Ito, N. Otomo, H. Suetsugu, S. Yoshino, M. Akiyama, K. Saito, Y. Ishikawa, C. Benner, P. Natarajan, P. T. Ellinor, T. Mushiroda, M. Horikoshi, M. Ikeda, N. Iwata, K. Matsuda, K. Matsuda, S. Niida, K. Ozaki, Y. Momozawa, S. Ikegawa, O. Takeuchi, K. Ito, C. Terao and P. Biobank Japan (2024). "Population-specific putative causal variants shape quantitative traits." Nature Genetics.
Kuderna, L. F. K., J. C. Ulirsch, S. Rashid, M. Ameen, L. Sundaram, G. Hickey, A. J. Cox, H. Gao, A. Kumar, F. Aguet, M. J. Christmas, H. Clawson, M. Haeussler, M. C. Janiak, M. Kuhlwilm, J. D. Orkin, T. Bataillon, S. Manu, A. Valenzuela, J. Bergman, M. Rouselle, F. E. Silva, L. Agueda, J. Blanc, M. Gut, D. de Vries, I. Goodhead, R. A. Harris, M. Raveendran, A. Jensen, I. S. Chuma, J. E. Horvath, C. Hvilsom, D. Juan, P. Frandsen, J. G. Schraiber, F. R. de Melo, F. Bertuol, H. Byrne, I. Sampaio, I. Farias, J. Valsecchi, M. Messias, M. N. F. da Silva, M. Trivedi, R. Rossi, T. Hrbek, N. Andriaholinirina, C. J. Rabarivola, A. Zaramody, C. J. Jolly, J. Phillips-Conroy, G. Wilkerson, C. Abee, J. H. Simmons, E. Fernandez-Duque, S. Kanthaswamy, F. Shiferaw, D. Wu, L. Zhou, Y. Shao, G. Zhang, J. D. Keyyu, S. Knauf, M. D. Le, E. Lizano, S. Merker, A. Navarro, T. Nadler, C. C. Khor, J. Lee, P. Tan, W. K. Lim, A. C. Kitchener, D. Zinner, I. Gut, A. D. Melin, K. Guschanski, M. H. Schierup, R. M. D. Beck, I. Karakikes, K. C. Wang, G. Umapathy, C. Roos, J. P. Boubli, A. Siepel, A. Kundaje, B. Paten, K. Lindblad-Toh, J. Rogers, T. Marques Bonet and K. K. Farh (2024). "Identification of constrained sequence elements across 239 primate genomes." Nature **625**(7996): 735-742.
Kumasaka, N., A. J. Knights and D. J. Gaffney (2016). "Fine-mapping cellular QTLs with RASQUAL and ATAC-seq." Nat Genet **48**(2): 206-213.
Leypold, N. A. and M. R. Speicher (2021). "Evolutionary conservation in noncoding genomic regions." Trends Genet **37**(10): 903-918.
Li, R. and G. Quon (2019). "scBFA: modeling detection patterns to mitigate technical noise in large-scale single-cell genomics data." Genome Biol **20**(1): 193.
Li, Y., Z. Li and F. Hu (2021). "Double-negative (DN) B cells: an under-recognized effector memory B cell subset in autoimmunity." Clin Exp Immunol **205**(2): 119-127.
Li, Y. K., H. Y. Wang, Y. Chen, X. Q. Shi, X. X. Zhang, K. Li, W. P. Fu and C. Sun (2023). "Single-nucleotide polymorphism rs6592645 confers asthma risk through regulating LRRC32 expression." Exp Ther Med **26**(3): 451.
Liu, X., S. Koyama, K. Tomizuka, S. Takata, Y. Ishikawa, S. Ito, S. Kosugi, K. Suzuki, K. Hikino, M. Koido, Y. Koike, M. Horikoshi, T. Gakuhari, S. Ikegawa, K. Matsuda, Y. Momozawa, K. Ito, Y. Kamatani and C. Terao (2024). "Decoding triancestral origins, archaic introgression, and natural selection in the Japanese population by whole-genome sequencing." Sci Adv **10**(16): eadi8419.
Mangan, R. J., F. C. Alsina, F. Mosti, J. E. Sotelo-Fonseca, D. A. Snellings, E. H. Au, J. Carvalho, L. Sathyan, G. D. Johnson, T. E. Reddy, D. L. Silver and C. B. Lowe (2022). "Adaptive sequence divergence forged new neurodevelopmental enhancers in humans." Cell **185**(24): 4587-4603 e4523.
Mantle, D., R. A. Heaton and I. P. Hargreaves (2021). "Coenzyme Q10 and Immune Function: An Overview." Antioxidants (Basel) **10**(5).
Martens, L. D., D. S. Fischer, V. A. Yepez, F. J. Theis and J. Gagneur (2024). "Modeling fragment counts improves single-cell ATAC-seq analysis." Nat Methods **21**(1): 28-31.
Maurano, M. T., R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds, R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin, S. Stehling-Sun, A. K. Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld, A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S. R. Sunyaev, R. Kaul and J. A. Stamatoyannopoulos (2012). "Systematic localization of common disease-associated variation in regulatory DNA." Science **337**(6099): 1190-1195.
Moser, T., K. Akgun, U. Proschmann, J. Sellner and T. Ziemssen (2020). "The role of TH17 cells in multiple sclerosis: Therapeutic implications." Autoimmun Rev **19**(10): 102647.

Mukherjee, R., P. Kanti Barman, P. Kumar Thatoi, R. Tripathy, B. Kumar Das and B. Ravindran (2015). "Non-Classical monocytes display inflammatory features: Validation in Sepsis and Systemic Lupus Erythematous." Sci Rep **5**: 13886.

Nait Saada, J., G. Kalantzis, D. Shyr, F. Cooper, M. Robinson, A. Gusev and P. F. Palamara (2020). "Identity-by-descent detection across 487,409 British samples reveals fine scale population structure and ultra-rare variant associations." Nat Commun **11**(1): 6130.

Nassir, N., R. Tambi, A. Bankapur, S. Al Heialy, N. Karuvantevida, H. H. Khansaheb, B. Zehra, G. Begum, R. A. Hameid, A. Ahmed, Z. Deesi, A. Alkhajeh, K. M. F. Uddin, H. Akter, S. A. Safizadeh Shabestari, O. Almidani, A. Islam, M. Gaudet, R. K. Kandasamy, T. Loney, A. A. Tayoun, N. Nowotny, M. Woodbury-Smith, P. Rahman, W. M. Kuebler, M. Yaseen Hachim, J. L. Casanova, B. K. Berdiev, A. Alsheikh-Ali and M. Uddin (2021). "Single-cell transcriptome identifies FCGR3B upregulated subtype of alveolar macrophages in patients with critical COVID-19." iScience **24**(9): 103030.

Nelson, R. K., A. Bush, J. Stokes, P. Nair and P. Akuthota (2020). "Eosinophilic Asthma." J Allergy Clin Immunol Pract **8**(2): 465-473.

O'Bleness, M., V. B. Searles, A. Varki, P. Gagneux and J. M. Sikela (2012). "Evolution of genetic and genomic features unique to the human lineage." Nat Rev Genet **13**(12): 853-866.

Oguchi, A., A. Suzuki, S. Komatsu, H. Yoshitomi, S. Bhagat, R. Son, R. J. P. Bonnal, S. Kojima, M. Koido, K. Takeuchi, K. Myouzen, G. Inoue, T. Hirai, H. Sano, Y. Takegami, A. Kanemaru, I. Yamaguchi, Y. Ishikawa, N. Tanaka, S. Hirabayashi, R. Konishi, S. Sekito, T. Inoue, J. Kere, S. Takeda, A. Takaori-Kondo, I. Endo, S. Kawaoka, H. Kawaji, K. Ishigaki, H. Ueno, Y. Hayashizaki, M. Pagani, P. Carninci, M. Yanagita, I. C. dagger, N. Parrish, C. Terao, K. Yamamoto, Y. Murakawa and I. C. Members (2024). "An atlas of transcribed enhancers across helper T cell diversity for decoding human diseases." Science **385**(6704): eadd8394.

Pollard, K. S., M. J. Hubisz, K. R. Rosenbloom and A. Siepel (2010). "Detection of nonneutral substitution rates on mammalian phylogenies." Genome Res **20**(1): 110-121.

Pollen, A. A., U. Kilik, C. B. Lowe and J. G. Camp (2023). "Human-specific genetics: new tools to explore the molecular and cellular basis of human evolution." Nat Rev Genet **24**(10): 687-711.

Radke, D. W., J. H. Sul, D. J. Balick, S. Akle, I. Alzheimer's Disease Neuroimaging, R. C. Green and S. R. Sunyaev (2021). "Purifying selection on noncoding deletions of human regulatory loci detected using their cellular pleiotropy." Genome Res **31**(6): 935-946.

Roadmap Epigenomics, C., A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y. C. Wu, A. R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shoresh, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K. H. Farh, S. Feizi, R. Karlic, A. R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthall, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. A. Boyer, P. L. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. Jones, W. Li, M. A. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L. H. Tsai, W. Wang, R. A. Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang and M. Kellis (2015). "Integrative analysis of 111 reference human epigenomes." Nature **518**(7539): 317-330.

Schaid, D. J., W. Chen and N. B. Larson (2018). "From genome-wide associations to candidate causal variants by statistical fine-mapping." Nat Rev Genet **19**(8): 491-504.

Silvert, M., L. Quintana-Murci and M. Rotival (2019). "Impact and Evolutionary Determinants of Neanderthal Introgression on Transcriptional and Post-Transcriptional Regulation." Am J Hum Genet **104**(6): 1241-1250.

Soskic, B., E. Cano-Gamez, D. J. Smyth, W. C. Rowan, N. Nakic, J. Esparza-Gordillo, L. Bossini-Castillo, D. F. Tough, C. G. C. Larminie, P. G. Bronson, D. Wille and G. Trynka (2019).

"Chromatin activity at GWAS loci identifies T cell states driving complex immune diseases." Nat Genet **51**(10): 1486-1493.
Stenson, P. D., E. V. Ball, M. Mort, A. D. Phillips, J. A. Shiel, N. S. Thomas, S. Abeysinghe, M. Krawczak and D. N. Cooper (2003). "Human Gene Mutation Database (HGMD): 2003 update." Hum Mutat **21**(6): 577-581.
Storey, J. D. (2002). "A Direct Approach to False Discovery Rates." Journal of the Royal Statistical Society Series B: Statistical Methodology **64**(3): 479-498.
Sullivan, P. F., J. R. S. Meadows, S. Gazal, B. N. Phan, X. Li, D. P. Genereux, M. X. Dong, M. Bianchi, G. Andrews, S. Sakthikumar, J. Nordin, A. Roy, M. J. Christmas, V. D. Marinescu, C. Wang, O. Wallerman, J. Xue, S. Yao, Q. Sun, J. Szatkiewicz, J. Wen, L. M. Huckins, A. Lawler, K. C. Keough, Z. Zheng, J. Zeng, N. R. Wray, Y. Li, J. Johnson, J. Chen, s. Zoonomia Consortium section, B. Paten, S. K. Reilly, G. M. Hughes, Z. Weng, K. S. Pollard, A. R. Pfenning, K. Forsberg-Nilsson, E. K. Karlsson and K. Lindblad-Toh (2023). "Leveraging base-pair mammalian constraint to understand genetic variation and human disease." Science **380**(6643): eabn2937.
Tanaka, N., M. Koido, A. Suzuki, N. Otomo, H. Suetsugu, Y. Kochi, K. Tomizuka, Y. Momozawa, Y. Kamatani, P. Biobank Japan, S. Ikegawa, K. Yamamoto and C. Terao (2021). "Eight novel susceptibility loci and putative causal variants in atopic dermatitis." J Allergy Clin Immunol **148**(5): 1293-1306.
Tashman, K. C., R. Cui, L. J. O'Connor, B. M. Neale and H. K. Finucane (2021). "Significance testing for small annotations in stratified LD-Score regression." medRxiv: 2021.2003.2013.21249938.
Terao, C., H. Yoshifuji, T. Matsumura, T. K. Naruse, T. Ishii, Y. Nakaoka, Y. Kirino, K. Matsuo, T. Origuchi, M. Shimizu, Y. Maejima, E. Amiya, N. Tamura, T. Kawaguchi, M. Takahashi, K. Setoh, K. Ohmura, R. Watanabe, T. Horita, T. Atsumi, M. Matsukura, T. Miyata, Y. Kochi, T. Suda, K. Tanemoto, A. Meguro, Y. Okada, A. Ogimoto, M. Yamamoto, H. Takahashi, S. Nakayamada, K. Saito, M. Kuwana, N. Mizuki, Y. Tabara, A. Ueda, I. Komuro, A. Kimura, M. Isobe, T. Mimori and F. Matsuda (2018). "Genetic determinants and an epistasis of LILRA3 and HLA-B*52 in Takayasu arteritis." Proc Natl Acad Sci U S A **115**(51): 13045-13050.
Thibodeau, A., A. Uyar, S. Khetan, M. L. Stitzel and D. Ucar (2018). "A neural network based model effectively predicts enhancers from clinical ATAC-seq samples." Sci Rep **8**(1): 16048.
Tran, D. Q., J. Andersson, R. Wang, H. Ramsey, D. Unutmaz and E. M. Shevach (2009). "GARP (LRRC32) is essential for the surface expression of latent TGF-beta on platelets and activated FOXP3+ regulatory T cells." Proc Natl Acad Sci U S A **106**(32): 13445-13450.
Trynka, G., C. Sandor, B. Han, H. Xu, B. E. Stranger, X. S. Liu and S. Raychaudhuri (2013). "Chromatin marks identify critical cell types for fine mapping complex trait variants." Nat Genet **45**(2): 124-130.
Visscher, P. M., N. R. Wray, Q. Zhang, P. Sklar, M. I. McCarthy, M. A. Brown and J. Yang (2017). "10 Years of GWAS Discovery: Biology, Function, and Translation." Am J Hum Genet **101**(1): 5-22.
Ward, L. D. and M. Kellis (2012). "Evidence of abundant purifying selection in humans for recently acquired regulatory functions." Science **337**(6102): 1675-1678.
Yan, F., D. R. Powell, D. J. Curtis and N. C. Wong (2020). "From reads to insight: a hitchhiker's guide to ATAC-seq data analysis." Genome Biol **21**(1): 22.
Yin, X., K. Kim, H. Suetsugu, S. Y. Bang, L. Wen, M. Koido, E. Ha, L. Liu, Y. Sakamoto, S. Jo, R. X. Leng, N. Otomo, V. Laurynenka, Y. C. Kwon, Y. Sheng, N. Sugano, M. Y. Hwang, W. Li, M. Mukai, K. Yoon, M. Cai, K. Ishigaki, W. T. Chung, H. Huang, D. Takahashi, S. S. Lee, M. Wang, K. Karino, S. C. Shim, X. Zheng, T. Miyamura, Y. M. Kang, D. Ye, J. Nakamura, C. H. Suh, Y. Tang, G. Motomura, Y. B. Park, H. Ding, T. Kuroda, J. Y. Choe, C. Li, H. Niiro, Y. Park, C. Shen, T. Miyamoto, G. Y. Ahn, W. Fei, T. Takeuchi, J. M. Shin, K. Li, Y. Kawaguchi, Y. K. Lee, Y. Wang, K. Amano, D. J. Park, W. Yang, Y. Tada, K. Yamaji, M. Shimizu, T. Atsumi, A. Suzuki, T. Sumida, Y. Okada, K. Matsuda, K. Matsuo, Y. Kochi, H. Japanese Research Committee on Idiopathic Osteonecrosis of the Femoral, L. C. Kottyan, M. T. Weirauch, S. Parameswaran, S.

Eswar, H. Salim, X. Chen, K. Yamamoto, J. B. Harley, K. Ohmura, T. H. Kim, S. Yang, T. Yamamoto, B. J. Kim, N. Shen, S. Ikegawa, H. S. Lee, X. Zhang, C. Terao, Y. Cui and S. C. Bae (2021). "Meta-analysis of 208370 East Asians identifies 113 susceptibility loci for systemic lupus erythematosus." Ann Rheum Dis **80**(5): 632-640.

Zeng, B., J. Bendl, C. Deng, D. Lee, R. Misir, S. M. Reach, S. P. Kleopoulos, P. Auluck, S. Marenco, D. A. Lewis, V. Haroutunian, N. Ahituv, J. F. Fullard, G. E. Hoffman and P. Roussos (2024). "Genetic regulation of cell type-specific chromatin accessibility shapes brain disease etiology." Science **384**(6698): eadh4265.

Zhang, J. Y., X. M. Wang, X. Xing, Z. Xu, C. Zhang, J. W. Song, X. Fan, P. Xia, J. L. Fu, S. Y. Wang, R. N. Xu, X. P. Dai, L. Shi, L. Huang, T. J. Jiang, M. Shi, Y. Zhang, A. Zumla, M. Maeurer, F. Bai and F. S. Wang (2020). "Single-cell landscape of immunological responses in patients with COVID-19." Nat Immunol **21**(9): 1107-1118.

Zhang, K., J. D. Hocker, M. Miller, X. Hou, J. Chiou, O. B. Poirion, Y. Qiu, Y. E. Li, K. J. Gaulton, A. Wang, S. Preissl and B. Ren (2021). "A single-cell atlas of chromatin accessibility in the human genome." Cell **184**(24): 5985-6001 e5919.

Zhou, J., C. L. Theesfeld, K. Yao, K. M. Chen, A. K. Wong and O. G. Troyanskaya (2018). "Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk." Nat Genet **50**(8): 1171-1179.

Zhou, J. and O. G. Troyanskaya (2015). "Predicting effects of noncoding variants with deep learning-based sequence model." Nat Methods **12**(10): 931-934.

Zoonomia, C. (2020). "A comparative genomics multitool for scientific discovery and conservation." Nature **587**(7833): 240-245.