Chromatin activity at GWAS loci identifies T cell states driving complex immune diseases

Blagoje Soskic1,2,9, Eddie Cano-Gamez1,2,9, Deborah J. Smyth1,2, Wendy C. Rowan3, Nikolina Nakic4, Jorge Esparza-Gordillo5, Lara Bossini-Castillo1, David F. Tough6, Christopher G. C. Larminie5, Paola G. Bronson7, David Willé8 and Gosia Trynka1,2*

Immune-disease-associated variants are enriched in active chromatin regions of T cells and macrophages. However, whether these variants function in specific cell states is unknown. Here we stimulated T cells and macrophages in the presence of 13 cytokines and profiled active and open chromatin regions. T cell activation induced major chromatin remodeling, while the presence of cytokines fine-tuned the magnitude of changes. We developed a statistical method that accounts for subtle changes in the chromatin landscape to identify SNP enrichment across cell states. Our results point towards the role of immune-disease-associated variants in early rather than late activation of memory CD4+ T cells, with modest differences across cytokines. Furthermore, variants associated with inflammatory bowel disease are enriched in type 1 T helper (TH1) cells, whereas variants associated with Alzheimer’s disease are enriched in different macrophage cell states. Our results represent an in-depth analysis of immune-disease-associated variants across a comprehensive panel of activation states of T cells and macrophages.

Functional interpretation of complex-disease-associated variants is challenging because the majority of loci mapped through genome-wide association studies (GWAS) reside in non-coding regions of the genome. Multiple studies have mapped GWAS variants to regulatory elements, such as open chromatin regions and regions tagged by histone modifications1–5, indicating the role of these elements in the regulation of gene expression. The functional effects of non-coding GWAS variants are difficult to deconvolute and may be specific to a particular cell type as well as a cell state, for example, the different stages of cell activation6. Integrating GWAS variants with cell-type-specific chromatin marks can provide insights into disease-causal cell types1,7. This approach has previously identified CD4+ T cells8 and monocytes9 as relevant cell types in the pathobiology of various complex immune diseases.

CD4+ T cells are key regulators of the immune response and are crucial in the protection against pathogens. One of the hallmarks of CD4+ T cells is their plasticity; in particular, the ability to differentiate into a range of cell states in response to environmental signals. CD4+ T cells undergo initial activation when they recognize antigens displayed by antigen-presenting cells in the context of co-stimulatory signals. Subsequently, activated T cells undergo proliferation and can be driven to differentiate into distinct T helper phenotypes, depending on the specific cytokines secreted by antigen-presenting cells. The major T helper types include Th1, Th2, Th17 and induced regulatory T cells, each of which has different functions in the immune response. Effector T helper phenotypes are defined by the specific cytokines that they secrete, which in turn instruct other immune cells to acquire different phenotypes. For example, the Th1 cytokine interferon-γ (IFNγ) polarizes macrophages to a proinflammatory (M1) phenotype with increased pathogen-killing ability, whereas the Th2 cytokine interleukin-4 (IL-4) induces a tissue-remodeling macrophage phenotype (M2)10. As such, the proper differentiation of T cells and macrophages after activation by cytokine signals is a crucial step in eliciting an appropriate immune response.

Although it is established that immune-disease-associated variants localize to chromatin regions that are specific to CD4+ T cells and monocytes, it is not yet known whether immune-disease-associated variants are further enriched in chromatin regions that are specific to a particular cytokine-induced cell state. To identify whether immune-disease-associated variants regulate cellular responses to cytokine polarization, we profiled chromatin accessibility using assay for transposase-accessible chromatin using sequencing (ATAC-seq), and active enhancers and promoters marked by histone H3 K27 acetylation (H3K27ac; see Methods) in naive and memory CD4+ T cells as well as macrophages across 55 cell activation states, including early and late responses to activation and cytokine polarization (Supplementary Table 1). We developed a statistical method for assessing SNP enrichment in chromatin marks to analyze the effects of immune-disease-associated variants in specific cell states.

**Results**

**Overview of the experimental design.** The enrichment of GWAS variants in CD4+ T cells places this cell type at the center of the dysregulated immune responses in disease pathobiology. Key steps in the regulation of the quality of an immune response include the initial activation and differentiation of CD4+ T cells and the subsequent interaction of polarized T cells with downstream effector cells such as macrophages, the activity of which is regulated by T-cell-derived factors. In this study, we focused on dissecting the role of immune-disease-associated risk variants in regulating this

---

1Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK. 2Open Targets, Wellcome Genome Campus, Cambridge, UK. 3Novel Human Genetics, GSK Research, GSK Medicines Research Centre, Stevenage, UK. 4Functional Genomics, Molecular Science and Technology R&D, GSK Medicines Research Centre, Stevenage, UK. 5Human Genetics, GSK Research, GSK Medicines Research Centre, Stevenage, UK. 6Epigenetics RU, Oncology R&D, GSK Medicines Research Centre, Stevenage, UK. 7Human Target Validation Core, RED Translational Biology, Biogen, Cambridge, MA, USA. 8Biostatistics, GSK Research, GSK Medicines Research Centre, Stevenage, UK. 9These authors contributed equally: Blagoje Soskic, Eddie Cano-Gamez. e-mail: gosia@sanger.ac.uk
circuitry. For this purpose, we stimulated monocyte-derived macrophages with T-cell-produced cytokines associated with inflammation and autoimmunity, including IFNγ, TNF, IL-4, IL-23 and IL-26 (Supplementary Table 1). As macrophages are part of the fast-responding innate immune system, we measured cytokine-induced activation at 6 h (early response) and 24 h (late response) and profiled the chromatin regulatory landscape. To mimic T cell activation in vitro, we stimulated T cells by delivering T cell receptor and CD28 signals using beads coated with anti-CD3 and anti-CD28 antibodies. In addition, we exposed cells to cytokine cocktails that promote the differentiation towards Tp1, Tp2, Tp17 or induced regulatory T cell fates, or to individual cytokines relevant to autoimmunity (IL-10, IL-21, IL-27, TNF and IFNβ)14–15 (Supplementary Table 1, see Methods). These stimuli were applied to memory and naive CD4+ T cells, which constitute the two major subsets of CD4+ T cells. We treated naive and memory cells separately because the two subsets differ in their response to stimulation16. Given that the response to stimulation develops over time17, we profiled T cells during both early and late activation. We defined early response as 16 h, in order to capture the regulation of gene expression before the first cell division. For the late response, we chose 5 d, which is when T cells acquire a defined effector phenotype. At each time point, we profiled the chromatin regulatory landscape by quantifying chromatin activity (H3K27ac ChiPmentation−seq (chromatin immunoprecipitation with sequencing-library preparation by Tn5 transposase)) and chromatin accessibility (ATAC-seq; Fig. 1a). We then integrated these chromatin profiles with disease-associated variants to identify the most disease-relevant cell states.

**Stimulation alters the chromatin landscape of immune cells.** We used H3K27ac and ATAC-seq reads to define active chromatin elements (peaks) in the profiled cell types. We detected an average of 25,613 (full range 13,466–38,939) H3K27ac peaks and 26,290 (full range 10,069–58,098) ATAC-seq peaks in T cells, and 34,708 (full range 27,605–42,771) H3K27ac peaks and 73,474 (full range 14,496–129,124) ATAC-seq peaks in macrophages (Supplementary Fig. 1b). We detected more peaks in T cells after stimulation compared to unstimulated T cells in the H3K27ac but not in the ATAC dataset (two tailed Student’s t-test, \( P = 0.0001 \) and \( P = 0.4 \) for H3K27ac and ATAC, respectively; Supplementary Fig. 1c) The identified peaks were highly reproducible across biological replicates (Supplementary Fig. 1d,e). On average, H3K27ac peaks were broader than ATAC-seq peaks (median size for H3K27ac and ATAC-seq peaks was 3,205 and 410 bp, respectively; Supplementary Fig. 2a). Furthermore, a larger proportion of ATAC-seq peaks were concentrated at transcriptional start sites (TSSs) compared to H3K27ac, indicating that ATAC-seq was more enriched in promoter regions (Supplementary Fig. 2bc). We also observed that ATAC-seq peaks were mostly tagged intronic, intergenic and promoter regions, whereas H3K27ac peaks were predominantly present in introns and intergenic regions (Supplementary Fig. 2c). Finally, most of H3K27ac peaks overlapped one or more ATAC-seq peaks (Supplementary Fig. 2d).

Taken together, we conclude that our maps of chromatin activity were of high quality and appropriately captured regulatory elements across the profiled cytokine-induced cell states.

We next investigated the effects of stimulation on the chromatin landscape of naive and memory T cells, as well as macrophages. In general, we observed that the main source of variation correlated with cell type, followed by stimulation (Supplementary Fig. 3), which induced thousands of peaks across all cell types (Fig. 1b). However, only a small proportion of peaks were specific to an individual cell state and the majority of peaks were shared between cell states (Fig. 1c); for example, in T cells, only 2% of H3K27ac and 0.8% of ATAC-seq peaks were specific to the condition. We then assessed the variability in levels of activity of each regulatory element across the cell states by quantifying the coefficient of variation of reads in peaks. We found that quantitative levels of chromatin activity were highly variable across different cytokine-induced cell states and both ATAC and H3K27ac peaks formed a continuous spectrum of variable peaks (Fig. 1c). The most variable peaks were in proximity to genes that are involved in cell differentiation, whereas the peaks with low variability were in proximity to genes involved in metabolic processes (Supplementary Fig. 2e). We also observed that the levels of chromatin accessibility were less variable than those of H3K27ac. For instance, loci that contained hallmark genes for cytokine-induced states showed marked differences in their H3K27ac profiles (Fig. 1d) but not in the ATAC-seq profile (Supplementary Fig. 4). This is in line with other reports, which have shown that chromatin activity marked by H3K27ac is more informative in discriminating between closely related cell states than is chromatin accessibility18,19.

**Accounting for peak properties refines SNP enrichment.** We next assessed whether immune-disease-associated variants were enriched in chromatin elements of any specific cell state in our data. Typically, disease-associated SNP enrichment analyses rely on the presence or absence of overlap between associated variants and regulatory regions20–23. However, in our dataset the majority of peaks were shared across cell states and therefore such a binary SNP−peak overlap approach would be unsuitable to discriminate between the different cellular conditions (Supplementary Fig. 5). A similar observation was previously made using partitioning heritability analysis of neuropsychiatric and metabolic disorders in highly correlated chromatin annotations of different brain regions24. To assess the immune disease enrichment across the different T cell and macrophage states, we developed a SNP enrichment method that we call CHEERS (Chromatin Element Enrichment Ranking by Specificity). In addition to SNP−peak overlap, our method takes into account peak properties as reflected by quantitative changes in read counts within peaks, corresponding to variable levels of H3K27ac or chromatin accessibility (Fig. 2, see Methods). In brief, we first construct a matrix of quantile-normalized read counts across peaks (Fig. 2a).

For each peak, we generate specificity scores, for which a high score is assigned to a peak with a higher read count in that cell state compared to all other states. Within each cell state, peaks are then ranked based on their specificity scores (specificity rank; Fig. 2b). To assess disease-associated SNP enrichment across the different cell states, for each locus we use the index variant and identify variants in strong linkage disequilibrium (LD; \( R^2 > 0.8 \)). Next, we identify peaks that overlap with the associated variants (Fig. 2c).

Notably, our method is peak-centric—that is, a peak that overlaps multiple disease-associated variants in a locus contributes to the final cell-type specificity score only once. Alternatively, if the associated variants within a locus intersect multiple peaks, each independent peak contributes to the final enrichment score. We then calculate the mean specificity rank of all peaks that intersect disease-associated variants and infer the significance of the observed SNP enrichment in the cell type by comparing it to a theoretical distribution (see Methods).

We used simulations to assess the sensitivity of CHEERS (see Methods). With more than 23% of variants overlapping the top 10% most cell-type-specific peaks, we observed over 80% power to detect a significant enrichment (\( P < 0.01 \)). However, at least 78% of SNPs that are needed to map within peaks ranked as low as 60–70% of cell-type-specific peaks to achieve the same power of detecting a significant enrichment (Supplementary Fig. 6). This indicates that CHEERS can detect a significant cell type enrichment only when a sufficient proportion of trait-associated variants overlap peaks with high cell-type specificity. On the basis of the results from these simulations, we concluded that CHEERS is able to identify enrichment in active chromatin elements across closely related cell types.
**Fig. 1** Quantitative changes in chromatin activity distinguish immune cell states. 

**a.** Overview of the study design. Naive and memory CD4+ T cells were stimulated with anti-CD3/anti-CD28 antibody-coated T cell activation beads. Macrophages were differentiated from monocytes by using macrophage colony-stimulating factor. All cell types were cultured in the presence of immune-disease-relevant cytokines and their chromatin activity was profiled at an early and a late time point. 

**b.** The number of differential H3K27ac and ATAC peaks after T cell activation with anti-CD3/anti-CD28 T cell activation beads and macrophage activation with TNF. iTreg, induced regulatory T cells. 

**c.** Proportion of activation-induced peaks that are shared within all T cell or macrophage states. Different colors represent the extent of sharing. Density plots show the coefficient of variation of the number of reads within ATAC and H3K27ac peaks. 

**d.** H3K27ac read pile-ups in proximity to condition-specific hallmark genes: CD3 (T cells), CD14 (macrophages), MKI67 (also known as Ki-67; T cell activation), TBX21 (T1), GATA3 (T2), IL35B (also known as EBI3; activated by induced regulatory T cells), IFIT2 (IFN-induced). Genomic coordinates (GRCh38) for each gene are labeled. H3K27ac tracks are generated after merging three biologically independent samples per cell state. CPM, counts per million.
CHEERS identifies cell types relevant for immune diseases. To further validate CHEERS, we used a catalog of 19 primary immune cell types assayed with H3K27ac ChIP–seq as a part of the BLUEPRINT project and tested the enrichment of SNPs associated with 12 complex diseases that have an immune component, including rheumatoid arthritis, allergies, asthma, celiac disease (CED), inflammatory bowel disease (IBD), Crohn’s disease, ulcerative colitis, multiple sclerosis, psoriasis, type 1 diabetes, systemic lupus erythematosus and systemic sclerosis. In agreement with previous results, we observed enrichment of rheumatoid arthritis-associated variants in chromatin regions that were active in memory CD4+ T and CD8+ T cells (Supplementary Fig. 7a). This demonstrated that CHEERS accurately detected enrichment in a dataset of diverse cell types. We observed that risk variants for allergies and asthma were enriched in chromatin regions that were specifically active in memory CD4+ and CD8+ T cells. Moreover, we found enrichment of risk variants for systemic lupus erythematosus and systemic sclerosis in chromatin elements that were specific to B cells (Supplementary Fig. 7a), as previously reported.

We found that the immune-mediated diseases that we tested fell into three groups (Supplementary Fig. 7b): diseases with a strong T cell component (asthma, allergies, IBD, rheumatoid arthritis, CED and type 1 diabetes), diseases with a strong B cell component (psoriasis, systemic lupus erythematosus and systemic sclerosis) and diseases with both B and T cell components (Crohn’s disease, multiple sclerosis and ulcerative colitis). Additionally, we observed that for systemic lupus erythematosus, CHEERS not only recapitulated the previously described enrichment in B cells, but also refined the enrichment further to class-switched B cells, with little signal in unswitched B cells (Supplementary Fig. 7a). On the basis of these results, we conclude that CHEERS is also suitable for the identification of enrichment of trait variants in active chromatin elements across related cell types.

Finally, to further validate CHEERS we used the Roadmap Epigenomics dataset, which contains H3K27ac data from a collection of 50 tissues and cell types, including a variety of non-immune tissues (https://egg2.wustl.edu/roadmap/data/byFileType/). We included three diseases in this analysis: multiple sclerosis, coronary artery disease and Parkinson’s disease (Supplementary Fig. 7d). We identified strong enrichment of GWAS loci for Parkinson’s disease in cell types from different brain regions and the adrenal gland. By contrast, GWAS loci for multiple sclerosis showed significant enrichment in immune cell types. This showed that CHEERS can correctly disentangle the biological differences between multiple sclerosis and Parkinson’s disease even though both affect the central nervous system. Moreover, we identified a significant enrichment of risk variants for coronary artery disease in chromatin regions that were specific to the aorta (top enrichment), lungs and multiple muscle tissues (both skeletal and smooth). Overall, these results demonstrate that CHEERS was able to identify relevant cell-type-specific enrichment of GWAS loci for both immune and non-immune traits.

Refining cell states relevant for immune diseases. We next tested whether GWAS variants associated with immune-mediated diseases were enriched in active chromatin regions that are specific to the cytokine-polarizing cell states in our dataset. In addition, we included GWAS variants associated with Alzheimer’s disease, as microglia—which are a subtype of macrophages—have been suggested to play a role in this disease. The tested traits included from 17 to 132 associated loci, resulting in a range of 25 (asthma) to

**Fig. 2 | Overview of the CHEERS method.** a, We first define a union of peak regions present across cell types and construct a matrix of normalized read counts. b, We then calculate and rank specificity scores. c, To test for enrichment of disease-associated SNPs, we take all of the index variants and identify variants in strong LD ($r^2 > 0.8$). We then identify peaks that overlap with the associated variants, calculate the mean specificity ranks and assess statistical significance.
317 (IBD) SNP–peak overlaps that contributed to the final enrich-
ment scores.

Variants associated with multiple sclerosis, rheumatoid arthri-
tis, CED, IBD, Crohn’s disease, ulcerative colitis, psoriasis, type
1 diabetes, allergy and asthma were predominantly enriched in
T cells, which is in line with the observations from the BLUEPRINT
data (Fig. 3). This enrichment was particularly strong in chroma-
tin regions that are specifically active during early activation of

---

**Fig. 3 | Enrichment of disease-associated SNPs in H3K27ac regions in cytokine-induced cell states.** One-sided $P$ values are reported from a discrete uniform distribution. CHEERS was performed after merging three biologically independent samples resulting in 127,723 peaks. GWAS data and the statistical methods are described in the Methods sections ‘GWAS data processing’ and ‘CHEERS’. The dotted line marks the nominal $P$-value threshold of 0.05; the solid gray line represents the Bonferroni-corrected significance threshold ($P \leq 9 \times 10^{-4}$). LDL, low-density lipoprotein. There are no data for IL-23-stimulated macrophages for 6 h, owing to the lack of biological material.
memory T cells (16h; Fig. 3). We did not observe significant differences between early and late activation of T cells, either in the number or the quality (that is, q-value) of H3K27ac peaks (Supplementary Fig. 1). Additionally, resting memory T cells—which were cultured without any stimulus—showed no significant enrichment in any of the tested diseases, suggesting that it is specifically the regulation of T cell activation that drives these signals. We did not observe T cell enrichment in control traits such as coronary artery disease and low-density lipoprotein cholesterol (Fig. 3). Finally, we found that enrichment was predominantly driven by peaks that were proximal to the TSS (Supplementary Fig. 8).

The observed enrichment of immune variants in early memory T cell activation was driven by a group of H3K27ac peaks that were highly specific to this cell state (Fig. 4a, Supplementary Fig. 9). For example, a group of GWAS variants for allergies overlapped a peak close to the chemokine (C-C motif) ligand 20 (CCL20) gene, which showed higher levels of acetylation specifically during early activation of memory T cells (Fig. 4b). Allergy-associated variants also mapped within a smaller group of peaks that were active specifically during early activation of naive T cells, explaining the secondary enrichment in activated naive T cells (Fig. 4a). A proportion of immune-disease-associated variants also mapped to peaks that were specific to early activation of both naive and memory CD4 T cells. For instance, rheumatoid arthritis-associated variants overlapped a peak near the acyl-CoA oxidase-like protein (ACOX1) gene, which is involved in T cell metabolism (Fig. 4b). Collectively, genes near peaks that drive the enrichment in early activation of memory T cells across all diseases were enriched in pathways such as T cell activation, T cell differentiation and leukocyte activation (Fig. 4d). This suggests that immune-disease-associated variants overlap enhancers and promoters that regulate gene expression programs during early activation of memory T cells.

Overall, we observed that individual cytokine conditions and late stages of T cell activation were enriched among a few selected T cell states and diseases. For example, multiple sclerosis-associated variants were significantly enriched after IL-27 stimulation in early memory T cell activation (P = 5.4 × 10^{-10}). This is concordant with previous studies that reported elevated levels of IL-27 in the cerebrospinal fluid of patients with multiple sclerosis [27]. Among the late activation states, naive T cells polarized with Th1 cytokines were of particular note as they were strongly enriched for IBD, Crohn’s disease, ulcerative colitis and CED (P = 9.92 × 10^{-14}, P = 1.11 × 10^{-10}, P = 4.89 × 10^{-14} and P = 1.01 × 10^{-10}, respectively). The H3K27ac peaks that intersect risk variants for these diseases with high specificity in Th1 cells were in close proximity to genes that are involved in Th1 biology, such as IFNγ and IL-12R (Fig. 4c). Furthermore, collectively these genes were enriched in hallmark pathways of Th1 cells such as IFNγ, IL-12 and JAK–STAT signaling (Fig. 4d).

Across all tested traits, only Alzheimer’s disease-associated variants were significantly enriched in macrophages (Fig. 3); however, we did not find any individual cytokine condition that was more significant than others. Genes in proximity to peaks that overlap Alzheimer’s disease-associated variants driving the macrophage signal were enriched in pathways, such as regulation of amyloid-β clearance and amyloid-β formation—processes that have been extensively studied in the context of Alzheimer’s disease pathology (Fig. 4d).

Finally, we observed that in the ATAC-seq data, the enrichment results were less significant (Supplementary Fig. 10a), although the global patterns of the most relevant disease-associated cell states replicated those of the H3K27ac data (Supplementary Fig. 10b). This agreed with our observations that ATAC-seq peaks were less variable. Similarly to H3K27ac data, ATAC-seq enrichments were mostly driven by the peaks that were proximal to a TSS (Supplementary Fig. 11).

In summary, our results suggest that variants associated with immune-mediated diseases play a role in genetic regulation during the early activation of CD4+ T cells, predominantly in memory T cells. We found that individual cytokine-induced cell states were enriched for variants associated with specific immune diseases. For example, risk variants for IBD, its two subphenotypes (ulcerative colitis and Crohn’s disease), as well as CED were enriched in chromatin elements specific to late activation of naïve T cells with Th1 polarizing cytokines. Our approach provides a framework for identifying disease relevant cell types and cell states and will inform the design of subsequent large-scale studies for functional interpretation of GWAS variants for complex immune diseases.

Discussion

Identifying the most relevant cellular context in which disease-associated variants function is critical for designing meaningful functional follow-up studies. Here, we used an in vitro system to identify cytokine-induced cell states that are relevant for the pathobiology of immune-mediated diseases. We found that most changes in the chromatin landscape of CD4+ T cells resulted from T cell receptor and CD28 stimulation alone, whereas the presence of specific cytokines had a modulatory effect on the levels of these changes. As a consequence, T cell states had very similar chromatin landscapes and currently available SNP enrichment methods could not distinguish between them. To address this, we developed CHEERS, a statistical method that takes into account quantitative changes in chromatin activity.

Using CHEERS, we were able to refine the enrichment of immune-disease-associated variants observed in previous studies [24,26] to specific cellular contexts. For 10 of the 12 immune diseases that we tested, we show that the associated variants were enriched during early activation of memory CD4+ T cells. A recent study, which profiled open chromatin regions across subsets of immune cells in resting and stimulated states, also identified enrichment of immune-disease-associated variants in T cell activation [25]. Our study now provides further evidence for the importance of this cell type in the biology of complex immune-mediated diseases and builds a case for dysregulation of specific cellular processes.

Immunophenotyping studies have identified elevated cytokine levels in patients with immune diseases and this has subsequently led to the development of disease-modifying therapies that target different cytokines [28–31]. This implies that effector T cells (corresponding to 5-d stimulations in our dataset) and the cytokines that they secrete drive autoimmune inflammation. By contrast, our results suggest that GWAS variants for immune-mediated diseases affect regulation mostly during the initial phase of memory T cell activation, but less so during the effector response seen at a later time point. Therefore, although cytokine-induced cell states could be key in the later stages of pathologic inflammation, they might not be triggering a disease. This emphasizes the importance of tight regulation of T cell activation, suggesting that many subtle effects of immune variants lead to dysregulated early cell responses. This agrees with observations from severe immune disorders in which, for instance, deficiency in the expression of one of the main regulators of T cell activation, CTLA-4, has been associated with the development of autoimmune diseases due to uncontrolled T cell activation [32–34]. Furthermore, it is worth noting that immune-mediated diseases are more often diagnosed in adults, which correlates with a shift in the frequency of T cells from naive to predominantly memory T cells. Therefore, our results suggest that focusing on the regulation of early activation of memory CD4+ T cell may provide an important axis for the development of new treatment options.

In addition to early activation of memory CD4+ T cells, we observed selected diseases for which the associated variants were enriched in distinctive cytokine-polarized cell states. For example,
IBD, Crohn’s disease, ulcerative colitis and CED risk variants were highly enriched in T cells polarized with T_{h}1 cytokines. This agrees with previous findings from immunophenotyping studies that have shown that lamina propria T cells express high levels of STAT4, T-bet and IL-12R, which are all induced by T_{h}1 polarizing cytokines. This could indicate that a proportion of IBD variants regulate the function of T_{h}1 cells, suggesting the involvement of this cell type in the development of IBD. Another example is asthma, for which we detected enrichment in naive T cells polarized towards the T_{h}2 phenotype.

Fig. 4 | Example loci that drive the enrichment of SNPs associated with immune diseases in cytokine-induced cell states. a, Loci that drive the enrichment of allergy-associated variants in early activation of memory T cells. Each row corresponds to a H3K27ac peak that intersects GWAS variants for allergy, whereas each column corresponds to a different cytokine-induced cell state. Shades of blue represent how specific each peak is to each cell state (specificity rank of the peak normalized to the mean specificity rank of all peaks). Different boxes highlight groups of peaks specific to different cell states. b, Read pile-ups at the selection of H3K27ac peaks that drive the enrichment of allergy-associated and rheumatoid arthritis-associated variants in early T cell activation. Genomic coordinates (GRCh38) for each gene are labeled. H3K27ac tracks were generated after merging three biologically independent samples per cell state. c, Read counts at a selection of H3K27ac peaks, driving the enrichment of IBD-associated variants in T_{h}1-stimulated T cells. Genomic coordinates (GRCh38) for each gene are labeled. H3K27ac tracks were generated after merging three biologically independent samples per cell state. d, Pathway enrichment analysis using all the genes in proximity to H3K27ac peaks driving the enrichment of GWAS variants in early T cell activation (all immune-mediated diseases), T_{h}1 cells (IBD and CED) and macrophages (Alzheimer’s disease). To determine these enrichments, we selected peaks with the top 10% specific peaks at the enriched cell states and assigned them to the closest gene. P values were calculated using gProfileR with default parameters. The number of genes used for enrichment analysis was 215 for memory CD4^{+} T cells (early activation), 27 for macrophages and 53 for CD4^{+} T cells polarized towards T_{h}1.
Across the tested immune diseases, we did not observe a significant enrichment in the macrophage stimulatory cell states, suggesting that autoimmune inflammation is mostly driven by T cells. By contrast, variants associated with Alzheimer’s disease point to a crucial role of macrophages in the disease. Recent studies have highlighted the role of microglia in Alzheimer’s disease\textsuperscript{22,23,37}, a type of resident macrophage in the central nervous system. Our results suggest that a proportion of Alzheimer’s disease-associated variants could be studied in an in vitro model of monocyte-derived macrophages. This could have important implications, as cells from the central nervous system are challenging to collect. Finally, the lack of stratification of Alzheimer’s disease-associated variants between the different cytokine-induced macrophage states could indicate that some Alzheimer’s disease-associated variants might regulate more general functions of the macrophage lineage. These functions could be closely linked to amyloid-\(\beta\) clearance, as suggested by our results.

Our study systematically profiles changes in the chromatin regulatory landscape that are induced by cytokines during activation of human immune cells. This provides a valuable resource to identify appropriate cell models for studying how genetic variants lead to diseases.

**Online content**

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41588-019-0493-9.

Received: 14 February 2019; Accepted: 30 July 2019; Published online: 23 September 2019

**References**

1. Maurano, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. *Nature* 337, 1190–1195 (2012).
2. ENCODE Project Consortium An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74 (2012).
3. Chen, L. et al. Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* 167, 1398–1414 (2016).
4. Trynka, G. et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat. Genet.* 45, 124–130 (2013).
5. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* 50, 621–629 (2018).
6. Fairfax, B. P. et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* 343, 1244949 (2014).
7. Trynka, G. et al. Disentangling the effects of colocalizing genomic annotations to functionally prioritize non-coding variants within complex trait loci. *Am. J. Hum. Genet.* 97, 139–152 (2015).
8. Farh, K. K.-H. et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 518, 337–343 (2015).
9. Alasco, K. et al. Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nat. Genet.* 50, 424–431 (2018).
10. Martinez, F. O. & Gordon, S. The M1 and M2 paradigm of macrophage activation: time for reassessment. *F1000Prime Rep.* 6, 13 (2014).
11. Khor, B., Gardet, A. & Xavier, R. J. Genetics and pathogenesis of inflammatory bowel disease. *Nature* 474, 307–317 (2011).
12. Yoshizaki, A. et al. Regulatory B cells control T-cell autoimmunity through IL-21-dependent cognate interactions. *Nature* 491, 264–268 (2012).
13. Mascanfroni, L. D. et al. IL-27 acts on DCs to suppress the T cell response and autoimmunity by inducing expression of the immunoregulatory molecule CD39. *Nat. Immunol.* 14, 1054–1063 (2013).
14. Feldmann, M. Development of anti-TNF therapy for rheumatoid arthritis. *Nat. Rev. Immunol.* 2, 364–371 (2002).
15. Hall, J. C. & Rosen, A. Type I interferons: crucial participants in disease amplification in autoimmunity. *Nat. Rev. Rheumatol.* 6, 40–49 (2010).
16. Glinos, D. A., Soskic, B., Jostins, L., Sansom, D. M. & Trynka, G. Genomic profiling of T cell activation reveals dependency of memory T cells on CD28 costimulation. Preprint at bioRxiv https://doi.org/10.1101/421099 (2018).
17. Ye, C. et al. Intersection of population variation and autoimmunity genetics in human T cell activation. *Science* 345, 1254665 (2014).
18. Minoux, M. et al. Gene bivalency at Polyclome domains regulates cranial neural crest positional identity. *Science* 355, eaal2913 (2017).
19. Mueller, B. et al. Widespread changes in nucleosome accessibility without changes in nucleosome occupancy during a rapid transcriptional induction. *Genes Dev.* 31, 451–462 (2017).
20. Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* 94, 559–573 (2014).
21. Hu, X. et al. Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. *Am. J. Hum. Genet.* 89, 496–506 (2011).
22. Calderon, D. et al. Inferring relevant cell types for complex traits by using single-cell gene expression. *Am. J. Hum. Genet.* 101, 686–699 (2017).
23. Tansey, K. E., Cameron, D. & Hill, M. J. Genetic risk for Alzheimer’s disease is concentrated in specific macrophage and microglial transcriptional networks. *Genome Med.* 10, 14 (2018).
24. Sénac, V. et al. Production of IL-27 in multiple sclerosis lesions by astrocytes and myeloid cells: modulation of local immune responses. *Glia* 64, 553–569 (2016).
25. Lalive, P. H. et al. Increased interleukin-27 cytokine expression in the central nervous system of multiple sclerosis patients. *J. Neuroinflammation* 14, 144 (2017).
26. Okada, Y. et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 506, 376–381 (2014).
27. Calderon, D. et al. Landscape of stimulation-responsive chromatin across diverse human immune cells. Preprint at bioRxiv https://doi.org/10.1101/409722 (2018).
28. Schett, G., Elewaut, D., McInnes, I. B., Dayer, J.-M. & Neurath, M. F. How cytokine networks fuel inflammation: toward a cytokine-based disease taxonomy. *Nat. Med.* 19, 822–824 (2013).
29. McInnes, I. B., Buckley, C. D. & Isaacs, J. D. Cytokines in rheumatoid arthritis — shaping the immunological landscape. *Nat. Rev. Rheumatol.* 12, 63–68 (2016).
30. Babaloo, Z. et al. Increased IL-17A but decreased IL-27 serum levels in patients with multiple sclerosis. *Iran. J. Immunol.* 10, 47–54 (2013).
31. Jamshidian, A., Shayanfard, A., Pourazar, A., Zarkesh-Esfahani, S.-H. & Ghargozloo, M. Biased Treg/Th17 balance away from regulatory toward inflammatory phenotype in relapsed multiple sclerosis and its correlation with severity of symptoms. *J. Neuroimmunol.* 262, 106–112 (2013).
32. Kuehn, H. S. et al. Immune dysregulation in human subjects with heterozygous germline mutations in CTLA4. *Science* 345, 1623–1627 (2014).
33. Lo, B. et al. Patients with LRBA deficiency show CTLA4 loss and immune dysregulation responsive to abatacept therapy. *Science* 349, 436–440 (2015).
34. Schuchert, D. et al. Autosomal dominant immune dysregulation syndrome in humans with CTLA4 mutations. *Nat. Med.* 20, 1410–1416 (2014).
35. Tivol, E. A. et al. Loss of CTLA-4 leads to massive lymphoproliferation and fatal multiorgan tissue destruction, revealing a critical negative regulatory role of CTLA-4. *Immunity* 3, 541–547 (1995).
36. Neurath, M. F. et al. The transcription factor T-bet regulates mucosal T cell activation in experimental colitis and Crohn’s disease. *J. Exp. Med.* 195, 1129–1143 (2002).
37. Gosselin, D. et al. An environment-dependent transcriptional network specifies human microglia identity. *Science* 356, eaal3222 (2017).

**Acknowledgements**

This work was funded by the Open Targets grant (OTAR040) awarded to G.T. G.T. is supported by the Wellcome Trust (grant WT206194). E.C.-G. is supported by a Gates Cambridge Scholarship (OPP11144). L.B.-C. is supported by the MRC Skills Development Fellowship (MR/N014995/1).

**Author contributions**

G.T. and B.S. conceived and designed the project. B.S., E.C.-G. and D.J.S. carried out the experimental work. B.S. and E.C.-G. performed the data analysis. B.S., E.C.-G., G.T., W.C.R., N.N., I.E.-G., D.E.T., C.G.C.L., D.W., L.B.-C. and P.G.B. interpreted the results. B.S., E.C.-G. and G.T. developed the CHEERS method. L.B.-C. calculated LD blocks. G.T. supervised the analysis. G.T., B.S., E.C.-G., W.C.R., N.N., I.E.-G., D.E.T., C.G.C.L., D.W. and P.G.B. wrote the manuscript.

**Competing interests**

The authors declare no competing interests.

**Additional information**

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0493-9.

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

Correspondence and requests for materials should be addressed to G.T.

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019.
Methods

Sample processing and cell activation. Blood samples were obtained from nine healthy individuals. Cellular conditions within each cell type (naïve CD4 T cells, memory CD4 T cells and monocyte-derived macrophages) were derived from three donors. The human biological samples were sourced ethically and their research use was in accordance with the terms of the informed consent under an IRB/EC-approved protocol (15/NW/0282). Peripheral blood mononuclear cells were isolated using Ficoll–Paque PLUS (GE Healthcare) density gradient centrifugation. Naïve and memory CD4 T cells were isolated from peripheral blood mononuclear cells using the EasySep naïve CD4 T cell isolation kit and memory CD4 T cell enrichment kit (StemCell Technologies) according to the manufacturer's instructions. T cell purity was above 94% in all samples (Supplementary Fig. 1a). T cells were stimulated with anti-CD3/anti-CD28 human T-activator Dynabeads (Invitrogen) at 1:2 bead:T cell ratio. Cytokines were added at the same time as the stimulus and cells were collected after 16 h and after 5 d (the list of cytokines and their concentrations are provided in Supplementary Table 1). Monocytes were isolated using the EasySep monocytes isolation kit according to the manufacturer's instructions. To generate macrophages, monocytes were plated in a 100 mm x 20 mm dish and cells were washed with 800 U ml⁻¹ macrophage colony-stimulating factor (PeproTech) for 7 d (Supplementary Fig. 1a). Following macrophage differentiation, cells were stimulated with cytokines for 6 h and 24 h (the list of cytokines and their concentrations are provided in Supplementary Table 1).

Flow cytometry. Cells were analyzed with an BD LSR Fortessa. All data obtained by flow cytometry were analyzed by Flowjo v.9.9.6 (TreeStar).

ChiPmentation-seq. To profile active enhancers and promoters, cells were washed with RPMI and Dynabeads were removed using a DynaMag magnet (Thermo Fisher). Next, the cells were resuspended at 1 x 10⁶ cells ml⁻¹ in FACS buffer (PBS buffer supplemented with 10% fetal calf serum and 1 mM EDTA) and chromatin was cross-linked by adding 1% formaldehyde and incubating at 37 °C for 5 min. To quench the reaction, we added glycine and washed the cells in cold PBS buffer. Cross-linked cell pellets were frozen in liquid nitrogen and stored at −80 °C until further processing.

To perform ChiPmentation, cross-linked cell pellets were thawed and processed using the iDeal ChiP-seq kit for histones (Diagenode) according to the manufacturer's instructions. In brief, cells were lysed and sonicated, and the resulting material was used for immunoprecipitation. Sonication was performed using a Bioruptor Pico (Diagenode), with six sound pulse cycles of 30 s each. For immunoprecipitation, we used ChiP-grade antibodies specific to human H3K27ac histones (Diagenode). A negative control that did not undergo immunoprecipitation (input) was also generated for each cell type. After immunoprecipitation, unbound material was removed by adding magnetic beads, as specified by the manufacturer's instructions. The bound chromatin was eluted by enzymatic digestion, purified and the resulting DNA was amplified by PCR.

Library preparation. To generate DNA libraries for sequencing, we used the Nextera DNA library preparation kit according to the manufacturer's instructions. In brief, DNA was amplified by PCR and fragments of inappropriate sizes were removed using Agencourt AMPure XP beads (BD). Finally, samples were pooled and loaded onto an Illumina HiSeq 2500 instrument for paired-end sequencing. To minimize batch effects, samples were randomized before sequencing. We obtained an average of 63 million paired-end reads per sample.

ATAC-seq. To profile open chromatin regions, stimulated cells were washed with RPMI and Dynabeads were removed using a DynaMag magnet (Thermo Fisher). Next, tagmentation was performed using the fast ATAC protocol. In brief, 50,000 cells were washed in cold PBS buffer and resuspended in 50 µl of Nextera tagmentation buffer supplemented with 0.01% digitonin and 2.5 µM Tris transposase (Nextera). Samples were then incubated at 37 °C and 800 r.p.m. for 30 min. After tagmentation, DNA was purified using MinElute PCR columns (Qiagen) according to the manufacturer's instructions and stored at −80 °C until library preparation.

Sequencing libraries were generated from tagmented DNA using the Nextera DNA library preparation kit according to the manufacturer's instructions. In brief, DNA was amplified by PCR and fragments of inappropriate sizes were removed using Agencourt AMPure XP beads (BD). Finally, samples were pooled and loaded onto an Illumina HiSeq 2500 instrument for paired-end sequencing. To minimize batch effects, samples were randomized before library preparation and before sequencing. We obtained an average of 58 million paired-end reads per sample.

ATAC-seq and ChiPmentation-seq analysis. We assessed the quality of the reads using fastqs and trimmed the adapters using skewer (v.0.2.2)⁴⁹. We then mapped reads to the human reference genome GRCh3M using bwa mem (v.0.7.3a)⁵⁰, only keeping uniquely mapped reads. We also removed PCR duplicates and mitochondrial reads from ATAC-seq data using sambtools (v.0.1.9)⁴⁰. This resulted in the final BAM files containing uniquely mapped, non-mitochondrial reads that were used for peak calling. Finally, we calculated insert size distributions using PICARD tools (v.2.6.0) to remove samples with over- or undersonicated chromatin, and samples with skewed distributions of insert sizes.

Peaks were called using MACS2 (v.2.1.1)⁵¹. For ATAC-seq, peaks were called using the standard MACS2 model and specifying --nomodel --shift -100 --extsize 200 on fragment BED files (both reads of a pair were merged into a single fragment, as the -f BAMPE option does not work when shift is specified). For H3K27ac, peak identification was performed on BAM files specifying -f BAMPE --broad --broad-cutoff 0.1 --nomodel --extsize 146. For each cell-type–stimulation pair we used an input from the same condition (for example, all memory 16H cells were sampled using input generated from memory 16H cells). We kept peaks that had values of q < 10⁻⁸ for H3K27ac data and q < 10⁻¹⁰ for ATAC-seq data. Finally, to ensure that the downstream analysis was carried out using high-quality data, we filtered out peaks that were present in less than two biological replicates within a condition (we refer to these as the confident peak set). The peak overlap between the assays was calculated using bedtools intersect (v.2.22.0), while the distance of a peak to the closest TSS and peak annotation was assessed using Homer's annotatePeaks.pl (v.4.10.3).

To assess the quality of our data, we calculated the fraction of reads in peaks (FRIP). We excluded 7 H3K27ac samples with FRIP < 10% and 15 ATAC-seq samples with FRIP < 5%. The average q values were 13.1 and 5.3 for ATAC-seq and ChiPmentation, respectively. The total number of samples remaining in our dataset after this quality control was 173 in the H3K27ac dataset and 168 in the ATAC-seq dataset. In addition, we removed samples with skewed distributions of insert sizes, or <10 million read pairs that passed quality control and <10,000 peaks were removed from the downstream analysis. Finally, we assessed peak reproducibility per cell state and kept only peaks that exist in at least two biological replicates of any condition. We also removed peaks on Y chromosome and alternative contigs. This generated a final set of H3K27ac and ATAC peaks that were used for GWAS enrichment analysis.

Additionally, two H3K27ac and three ATAC-seq samples that did not cluster with the corresponding group in principal component analysis were discarded from further analysis. After quality-control filtering, we kept a total of 173 H3K27ac and 168 ATAC-seq samples.

To define differentially accessible regions and differentially modified histone regions, we used DESeq². We compared all conditions to the resting state or to TH0 and used a Benjamini–Hochberg controlled false discovery rate of 10% and an absolute fold change ≥ 1.

BLUEPRINT ChiP-seq analysis. Raw data for H3K27ac ChiP-seq of 19 immune cell types were obtained from the BLUEPRINT consortium. These files were processed following the same data analysis pipeline as described above.

GWAS data processing. GWAS variants associated with 16 complex diseases at the genome wide significance threshold were obtained from previous studies. We included in our analysis: allergies⁴⁵, Alzheimer’s disease⁴⁶, asthma⁴⁷, CED⁴⁸, coronary artery disease⁴⁹, levels of low-density lipoprotein⁵⁰, Crohn’s disease⁵¹, IBD⁵², multiple sclerosis⁵³, Parkinson’s disease⁵⁴, psoriasis⁵⁵, rheumatoid arthritis⁵⁶, systemic lupus erythematosus⁵⁷, systemic sclerosis⁵⁸, type 1 diabetes⁵⁹ and ulcerative colitis⁶⁰. Variants mapping to the HLA region were removed. SNPs were finally pruned by LD using a 50-kb window and R² < 0.1. Variants with a minor allele frequency that was less than 5% were excluded. The final list of variants was used as an input for SNP enrichment analysis.

Disease enrichment with GoShifter. We ran GoShifter on ATAC and H3K27ac peaks as described previously. We ran 10,000 permutations. Command used: python goshifter.py -a SNP_list -a annotation/file/path -p 10000 -L D/file/path -o output_name.

CHEERS. Quality-controlled BAM files corresponding to biological replicates of the same condition were merged using samtools. Then, for each cell type and cell state, we quantified the number of reads within peak i in cell state j as:

\[ Z_{ij} = C_i \times \max \left( \frac{\sum C_{ij}}{\sum C_i} \right) \]

where C is the number of reads falling within peak i in cell state j.

To ensure that our analysis was not confounded by the low confidence peaks, we removed the bottom tenth percentile of peaks with the lowest read counts and obtained a final count of 127,723 peaks for H3K27ac ChiPmentation and 136,692 for ATAC-seq.

To compare the peaks across the cell types and cell conditions, we also quantile normalized the library size-corrected peak counts. We then transformed the read count of each peak into a score that reflects cell-type specificity (specificity score).

For the specificity score, we divided the normalized read counts of each peak
in each condition by the Euclidean norm for that peak across all conditions, as described in the following formula:

\[ S_{ij} = \frac{Z_{ij}}{\sqrt{\sum_{r=1}^{N_x} p_{ij}^r}} \]

where \( S_{ij} \) is the specificity score and \( Z_{ij} \) is the normalized number of reads within peak \( j \) in condition \( i \). This score is a number from 0 to 1, where 1 means that a peak has high read counts in only one cell state and 0 means the peak shows no read in that cell state.

Then, within each cell state, peaks were ranked by specificity score from the highest to the lowest score, for which the peak that was the least specific to the cell state was ranked 1. If multiple peaks have equal specificity scores, the same lower rank is assigned to all of them.

To test for disease SNP enrichment, we take all of the index variants, identify variants in LD (\( r^2 > 0.8 \)) and overlap them with peaks. Notably, our method is peak-centric, that is, a peak that overlaps multiple disease-associated variants in a locus contributes to the final cell-type specificity score only once. On the other hand, if within a locus the associated variants intersect multiplepeaks, each independent peak contributes to the final enrichment score. We then calculate the mean specificity rank of all peaks that intersect disease-associated variants. We inferred the significance of the observed enrichment by fitting a discrete uniform distribution. Within each cell state, all ranks (1, 2, 3, …, \( N \)) can be observed with an equal probability and thus follow a discrete uniform distribution (Kolmogorov–Smirnov test, \( P=1 \)) with mean (\( \mu \)):

\[ \mu = \frac{1 + N}{2} \]

and variance:

\[ \sigma^2 = \frac{\text{max} - \text{min} + 1)^2}{12} - 1 \]

where max and min are the maximal and minimal ranks in the dataset. When we substitute max with number of peaks (\( N \)) and min with the minimal rank (1), we obtain the following formula:

\[ \sigma^2 = \frac{N^2 - 1}{12} \]

which, under the central limit theorem converges to:

\[ \sigma^2 = \frac{N^2 - 1}{12\sigma} \]

where \( \sigma^2 \) is the variance of the mean of the \( n \) peak ranks overlapping the GWAS SNPs, assuming that these overlap at random.

Finally, we calculate \( P \) values as:

\[ P = 1 - \Phi \left( \frac{x - \mu}{\sigma} \right) \]

where the \( \Phi \) function is a cumulative normal distribution for \( x \sim N(\mu, \sigma^2) \); \( x \) is the observed rank, and \( \mu \) and \( \sigma \) are the expected values under the null hypothesis that all ranks occurred as random.

To ensure that this method accounts for any possible unidentified properties of the data, such as correlations, we also assessed the significance of the enrichment using an empirical, permutation-based strategy. For this strategy, within each of the tested cell states, we randomly sampled sets of peaks, matching to the number of peaks that overlapped GWAS variants, and calculated the mean of their ranks. We repeated this process \( N \) times, assessed the frequency at which the mean of the permuted ranks was greater or equal to the mean of the observed ranks, and derived an empirical \( P \) value:

\[ P = \frac{\text{num}(\text{permuted} > \text{observed})}{N} \]

where \( N \) is the number of permutations. Both approaches yielded similar results although the \( P \) values from CHEERS were not limited by the number of permutations (\( R^2 = 0.96, P < 2.2 \times 10^{-16} \)).

Power calculations for CHEERS. To estimate the power of CHEERS, we simulated 100 SNP–peak overlaps and tested for enrichment. Each simulation was repeated 100\( \times \) and power was estimated as the percentage of simulations that yielded a significant enrichment (\( P < 0.01 \)). We designed the simulations such that a given percentage of SNPs (ranging from 0% to 100%), but not the remaining SNPs, always overlapped with peaks in the top tenth percentile of specificity. Finally, to test how specific the peaks needed to be for our method to detect enrichment, we repeated the simulations at lower specificity percentiles.

Pathway enrichment analysis. Pathway enrichment analysis and statistics were performed using g:Profiler (v. 0.6.1).

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

Data availability
All of the raw-sequencing data are deposited in the EGA. ATAC-seq data are available at EGAS00001003501. H3K27ac data are available at EGAS00001002749.

Code availability
CHEERS code is available through GitHub (https://github.com/trynkalaab/CHEERS).

References
48. Schmitt, C., Rendeiro, A. F., Sheffield, N. C. & Bock, C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat. Methods 12, 963–965 (2015).
49. Corces, M. R. et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat. Genet. 48, 1193–1203 (2016).
50. Jiang, H., Lei, R., Ding, S.-W. & Zhu, S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics 15, 182 (2014).
51. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 26, 589–595 (2010).
52. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).
53. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137 (2008).
54. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).
55. Ferreira, M. A. et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. Nat. Genet. 49, 1752–1757 (2017).
56. Jansen, I. E. et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk. Nat. Genet. 51, 404–413 (2019).
57. Demenais, F. et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. Nat. Genet. 50, 42–53 (2018).
58. Trynka, G. et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat. Genet. 43, 1193–1201 (2011).
59. Nelson, C. P. et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. Nat. Genet. 49, 1385–1391 (2017).
60. Hoffmann, T. J. et al. A large electronic-health-record-based genome-wide study of serum lipids. Nat. Genet. 50, 401–413 (2018).
61. Jositsch, L. et al. Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 491, 119–124 (2012).
62. International Multiple Sclerosis Genetics Consortium (IMSGC) Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nat. Genet. 45, 1353–1360 (2013).
63. Chang, D. et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson’s disease risk loci. Nat. Genet. 49, 1511–1516 (2017).
64. Tiao, L. C. et al. Large scale meta-analysis characterizes genetic architecture for common psoriasis associated variants. Nat. Commun. 8, 15382 (2017).
65. Langefeld, C. D. et al. Transancestral mapping and genetic load in systemic lupus erythematosus. Nat. Commun. 8, 16021 (2017).
66. Bossini-Castillo, L., López-Isac, E., Mayes, M. D. & Martin, J. Genetics of systemic sclerosis. Semin. Immunopathol. 37, 443–451 (2015).
67. Onengut-Gumuscu, S. et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. Nat. Genet. 47, 381–386 (2015).
68. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930 (2014).
Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection
BLUEPRINT data was downloaded from EGA (http://ega-archive.org). Roadmap data were downloaded from: https://egg2.wustl.edu/roadmap/data/byFileType/

Data analysis
All codes were written in publicly available version of Python 2.7 and plots were made using R 3.4.2. GoShifter code was downloaded from github (https://github.com/immunogenomics/goshifter/). CHEERS code is available through github (https://github.com/trynkaLab/CHEERS). We trimmed the adapters using skewer (v0.2.2). We mapped reads to the human genome reference GRCh38 using bwa mem (v0.7.9a). PCR duplicates and mitochondrial reads were removed with samtools (v0.1.9). We calculated insert size distributions using PICARD tools (v2.6.0). Peaks were called using MACS2 (v2.2.1). Peaks were merged using bedtools (v2.22.0) and the number of reads within peaks was quantified using featureCounts (v1.5.1). Peaks were annotated using annotatePeaks.pl (v4.10.3). To define differentially accessible regions and differentially modified histone regions, we used DESeq2 (v1.18.1). Flow cytometry data were analyzed with FlowJo (v9.9.6).

Data

Policy information about availability of data

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All the raw sequence files are deposited in EGA. ATAC-seq data: https://www.ebi.ac.uk/ega/studies/EGAS00001003501; H3K27ac data: https://www.ebi.ac.uk/ega/studies/EGAS00001002749
Field-specific reporting

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

☑️ Life sciences

☑️ Behavioural & social sciences

☐ Ecological, evolutionary & environmental sciences

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

Life sciences study design

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

| Sample size | We performed ATAC-seq and H3K27ac ChM-seq on 55 cell states. Each condition was done on 3 donors that were used as replicates of each other. The sample size is consistent with the number of donors used by international consortia for chromatin profiling (eg. BLUEPRINT, ROADMAP). |
| Data exclusions | To ensure that the downstream analysis was carried out using high quality data, we excluded H3K27ac samples with FRiP less than 10% and ATAC-seq samples with FRiP below 5%. We kept peaks that had q value < 10^-2 in H3K27ac and q value < 10^-4 in ATAC-seq. All criteria were established based on the distribution within our data. We also filtered out peaks present in less than two biological replicates within a condition to decrease the chances of spurious disease enrichments. |
| Replication | To assure that the findings of CHEERS were accurate we performed permutations which resulted to the same conclusions. We also applied CHEERS on BLUEPRINT data and recapitulated previous findings. |
| Randomization | We applied randomization to library preparation, the experiment set up and pooling samples for sequencing to reduce batch effects. |
| Blinding | Not applicable. Human blood used in this study came from Cambridge NHS Blood and Transplantation (UK national health service) and donors are not identifiable. |

Reporting for specific materials, systems and methods

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

| Materials & experimental systems | Methods |
| n/a | Involved in the study |
| ☑️ | Antibodies |
| ☑️ | Eukaryotic cell lines |
| ☑️ | Palaeontology |
| ☑️ | Animals and other organisms |
| ☑️ | Human research participants |
| ☑️ | Clinical data |
| ☑️ | ChIP-seq |
| ☑️ | Flow cytometry |
| ☑️ | MRI-based neuroimaging |

Antibodies

| Antibodies used | H3K27ac polyclonal antibody (Diagenode, Catalog No: C15410174, Lot no: A.7071-001P, 1.2ug/ul) |
| Antibodies used | Anti-human CD4, APC (BioLegend; Clone: OKT4, Catalog No: 317416, Lot no: B217365, Dilution: 1:100) |
| Antibodies used | Anti-human CD45RA, Brilliant Violet 785 (BioLegend; Clone: HI100, Catalog No: 304140, Lot no: B246562, Dilution: 1:100) |
| Antibodies used | Anti-human CD45RO, PE-Cyanine7 (BioLegend; Clone: UCHL1, Catalog No: 304229, Lot no: B219490, Dilution: 1:100) |
| Antibodies used | Anti-human CD14 AlexaFluor488 (BioLegend, Clone: MSE2, Catalog No: 301811, Lot no: B238087, Dilution: 1:100) |
| Antibodies used | Anti-human CD68 PE/Cy7 (BioLegend, Clone Y1/82A, Catalog No: 333816, Lot no: B192642, Dilution: 1:100) |
| Antibodies used | Anti-human MerTK-BV421 (BioLegend, Clone: 590H11G1E3, Catalog No: 367602, Lot no: B242529, Dilution: 1:100) |
| Antibodies used | Anti-CD3/anti-CD28 human T-Activator Dynabeads® (ThermoFisher, Catalog No: 111.31D, Lot no: 00670465, Concentration: 1 bead per 2 Tcells) |

Validation

All flow cytometry antibodies that we used for purity check were validated for detecting human proteins by the manufacturer. we performed qPCR on chromatin pull-downs to validate H3K27ac Ab
Human research participants

Policy information about studies involving human research participants

Population characteristics  
British Caucasians. 27-60 years old (median: 54), 8 males and 1 female.

Recruitment  
Blood samples were obtained from nine healthy individuals, three individuals contributed to each of the cell types in the experiments, i.e. CD4 naive T cells, CD4 memory T cells, and monocyte derived macrophages. All participants are blood donors from Cambridge NHS Blood and Transplantation (UK national health service). This cohort is general healthy population, male biased. We do not think that these biases impact the results.

Ethics oversight  
Ethics were obtained from Wellcome Sanger Institute and the National Research Ethics Committee. Note that full information on the approval of the study protocol must also be provided in the manuscript.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links

May remain private before publication.

https://wwwdev.ebi.ac.uk/ega/studies/EGAS00001002749

Files in database submission

There are 183 ATAC-seq files in: https://wwwdev.ebi.ac.uk/ega/studies/EGAS00001003501

There are 192 H3K27ac ChM-seq files in: https://wwwdev.ebi.ac.uk/ega/studies/EGAS00001002749

Genome browser session

(e.g. UCSC)

no longer applicable

Methodology

Replicates  
Each cellular condition was performed in 3 biological replicates.

Sequencing depth

1) ATAC-seq: Total number of reads: 27.9-144.3M (median: 59.2M); Uniquely mapped reads: 26.3-134.3 (median: 55.4)

2) H3K27ac ChM-seq: Total number of reads: 24.2M-166.9M (median: 63.7M); Uniquely mapped reads: 21.96-94.18 (median: 25.6)

Antibodies  
H3K27ac polyclonal antibody, Diagenode, Catalog No: C15410174, Lot no: A.7071-001P , 1.2ug/ul

Peak calling parameters

For ATAC-seq, peaks were called using the standard MACS2 model and specifying --nomodel --shift -100 --extsize 200 on fragment BED files (both reads of a pair were merged into a single fragment, since -f BAMPE option does not work when shift is specified). For H3K27ac ChM peaks were called on BAM files specifying -f BAMPE --broad --broad-cutoff 0.1 --nomodel --extsize 146.

Data quality

We kept peaks that had q value < 10e-2 in H3K27ac and q value < 10e-4 in ATAC-seq. Finally, to ensure that the downstream analysis was carried out using high quality data, we filtered out peaks present in less than two biological replicates within a condition (we refer to these as the confident peak set).

Software

We trimmed the adapters using skewer (v0.2.2). We mapped reads to the human genome reference GRCh38 using bwa mem (v0.7.9a). PCR duplicates and mitochondrial reads were removed with samtools (v0.1.9). We calculated insert size distributions using PICARD tools (v2.6.0). Peaks were called using MACS2 (v2.1.1). Peaks were merged using bedtools (v2.22.0) and the number of reads within peaks was quantified using featureCounts (v1.5.1). To define differentially accessible regions and differentially modified histone regions, we used DESeq2 (v1.18.1). Peaks were annotated using annotatePeaks.pl (v4.10.3). CHEERS code is available through github (https://github.com/trynkaLab/CHEERS).

Flow Cytometry

Plots

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.
### Methodology

#### Sample preparation

Blood samples were obtained from nine healthy individuals, sequencing data for all the conditions within a cell type were generated from three individuals. The human biological samples were sourced ethically and their research use was in accord with the terms of the informed consents under an IRB/EC approved protocol (15/NW/0282). Peripheral blood mononuclear cells (PBMCs) were isolated using Ficoll-Paque PLUS (GE healthcare, Buckingham, UK) density gradient centrifugation. Naive and memory CD4+ T cells were isolated from PBMCs using EasySep® naive CD4+ T cell isolation kit and memory CD4+ T cell enrichment kit (StemCell Technologies, Meylan, France) according to the manufacturer’s instructions. T cells were surface stained with CD4-APC as well as CD45RA-BV785 or CD45RO-PE. Monocytes were isolated using EasySep® monocytes isolation kit according to the manufacturer’s instructions. In order to generate macrophages, monocytes were plated in a 100mm x 20mm dish and cells were treated with 800 U/ml M-CSF (PeproTech) for seven days. Macrophages were resuspended in n 90μl 1xPerm buffer (BD Cytofix/Cytoperm kit (BD 554714)) containing 5μl Fc-block. Following 10min incubation a cocktail containing CD14-PE, CD68-PE/Cy7, MerTK-BV421 and HLADR-FITC was added to the samples and they were stained on ice for 30min.

#### Instrument

- BD LSR Fortessa

#### Software

- FlowJo Ver 9.9.6

#### Cell population abundance

Cells were isolated with isolation kits as described above

#### Gating strategy

We gated on FSC-A/SSc-A for cells and FSC-A/FSC-W for singlets which was proceeded by the gating for purity markers. FACS analysis was performed for purity checks, the cells that undergone sequencing were isolated using magnetic beads.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.