Published in final edited form as:
Nat Genet. 2019 October 01; 51(10): 1486–1493. doi:10.1038/s41588-019-0493-9.

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

Blagoje Soskic[#1,2], **Eddie Cano-Gamez**[#1,2], **Deborah J. Smyth**[1,2], **Wendy C. Rowan**[3], **Nikolina Nakic**[4], **Jorge Esparza-Gordillo**[5], **Lara Bossini-Castillo**[1], **David F. Tough**[6], **Christopher GC Larminie**[5], **Paola G. Bronson**[7], **David Willé**[8], **Gosia Trynka**[1,2,*]

[1]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton CB10 1SA, UK

[2]Open Targets, Wellcome Genome Campus, Cambridge CB10 1SA, UK

[3]Novel Human Genetics, GSK Research, GSK Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, UK

[4]Functional Genomics, Molecular Science and Technology R&D, GSK Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, UK

[5]Human Genetics, GSK Research, GSK Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, UK

[6]Epigenetics RU, Oncology R&D, GSK Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, UK

[7]Human Target Validation Core, RED Translational Biology, Biogen, Cambridge, MA 02142, USA

[8]Biostatistics, GSK Research, GSK Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, UK

[#] These authors contributed equally to this work.

Abstract

Immune disease variants are enriched in active chromatin regions of T cells and macrophages. However, whether these variants function in specific cell states is unknown. We stimulated T cells and macrophages in the presence of thirteen cytokines and profiled active and open chromatin regions. T cell activation induced major chromatin remodeling, while cytokines fine-tuned the

[*]Correspondence to: Gosia Trynka (gosia@sanger.ac.uk).

Data Availability Statement

CHEERS code is available through github (https://github.com/trynkaLab/CHEERS) 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.

Author Contributions

GT and BS conceived and designed the project. BS, ECG and DJS carried out the experimental work. BS and ECG performed the data analysis. BS, ECG, GT,WCR, NN, JEG, DT, CL, DW, LBC and PGB interpreted the results. BS, DW, ECG and GT developed the *CHEERS* method. LBC calculated LD blocks. GT supervised the analysis. GT, BS, ECG, WR, NN, JEG, DT, CL, DW and PGB wrote the manuscript.

Competing Interests Statement

All authors declare no competing interests.

Europe PMC Funders Author Manuscripts

magnitude of changes. We developed a statistical method that accounts for subtle changes in chromatin landscape to identify SNP enrichment across cell states. Our results point towards the role of immune disease variants in early rather than late activation of memory CD4+ T cells, with modest differences across cytokines. Furthermore, inflammatory bowel disease variants are enriched in Th1 cells while Alzheimer's disease variants are enriched in different macrophage cell states. Our results represent an in-depth analysis of immune disease variants across a comprehensive panel of activation states of T cells and macrophages.

Introduction

Functional interpretation of complex disease 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 modifications[1–5], implicating their role in gene expression regulation. The functional impact of non-coding GWAS variants is difficult to deconvolute and may be specific to a particular cell type as well as cell state context, such as different stages of cell activation[6]. Integrating GWAS variants with cell type specific chromatin marks can provide insights into disease causal cell types[1,4,7]. This approach has previously identified CD4+ T cells[4,8] and monocytes[6,9] as relevant cell types in the pathobiology of various complex immune diseases.

CD4+ T cells are key regulators of 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 antigen displayed by antigen-presenting cells (APCs) in the context of co-stimulatory signals. Subsequently, activated T cells undergo proliferation and can be driven to differentiate into distinct T helper (Th) phenotypes, depending on the specific cytokines secreted by APCs. The major Th types include Th1, Th2, Th17 and induced regulatory T cells (iTregs), each exerting different functions in the immune response. Effector Th 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 IFN-γ polarizes macrophages to a proinflammatory (M1) phenotype with increased pathogen killing ability, while the Th2 cytokine IL-4 induces a tissue remodeling macrophage phenotype (M2)[10]. As such, the proper differentiation of T cells and macrophages following cytokine signals is a crucial step in eliciting an appropriate immune response.

Although it is established that immune disease variants localize to chromatin regions specific to CD4+ T cells and monocytes, it is not yet known if immune disease variants are further enriched in chromatin regions specific for a particular cytokine-induced cell state. To identify whether immune disease variants regulate cellular responses to cytokine polarization, we profiled chromatin accessibility using ATAC-seq, and active enhancers and promoters marked by H3K27ac (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 new statistical method

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

for assessing SNP enrichment in chromatin marks to point towards the effects of immune disease variants in specific cell states.

Results

Overview of the experimental design

The GWAS link to CD4+ T cells places this cell type at the heart of dysregulated immune responses in disease pathobiology. Key steps in regulating 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, whose activity is regulated by T cell-derived factors. In this study we focused on dissecting the role of immune disease risk variants in regulating this circuitry. For this purpose, we stimulated monocyte-derived macrophages with T-cell-produced cytokines associated with inflammation and autoimmunity, including IFNγ, TNFα, IL-4, IL-23 and IL-26 (Supplementary Table 1). Since macrophages are part of the fast-responding innate immune system, we measured cytokine induced activation at six hours (early response) and 24 hours (late response) and profiled the chromatin regulatory landscape. To mimic T cell activation *in vitro,* we stimulated T cells by delivering T cell receptor (TCR) and CD28 signals using beads coated with anti-CD3 and anti-CD28 antibodies. In addition, we exposed cells to cytokine cocktails promoting differentiation towards Th1, Th2, Th17 or iTreg cell fates, or to individual cytokines relevant to autoimmunity (IL-10, IL-21, IL-27, TNFα, and IFNβ)[11–15] (Supplementary Table 1, 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 stimulation[16]. Given that the response to stimulation develops over time[17], we profiled T cells during both early and late activation. We defined early response as 16 hours in order to capture gene expression regulation prior to the first cell division. For late response we chose five days, 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) and chromatin accessibility (ATAC-seq; Figure 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 (13,466 - 38,939) H3K27ac peaks and 26,290 (10,069 - 58,098) ATAC-seq peaks in T cells, and 34,708 (27,605 - 42,771) H3K27ac peaks and 73,474 (14,496 - 129,124) ATAC-seq peaks in macrophages (Supplementary Figure 1b). We detected more peaks in T cells after stimulation compared to unstimulated T cells in H3K27ac but not in the ATAC (two tailed t-test $p = 0.0001$ and $p = 0.4$ for H3K27ac and ATAC, respectively) (Supplementary Figure 1c). The identified peaks were highly reproducible across biological replicates (Supplementary Figure 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 Figure 2a). Furthermore, a larger proportion of ATAC-seq peaks concentrated at transcriptional start sites (TSS)

compared to H3K27ac, reflecting that ATAC-seq was more enriched in promoter regions (Supplementary Figure 2b,c). We also observed that, ATAC-seq peaks tagged mostly intronic, intergenic and promoter regions, while H3K27ac peaks were present in introns and intergenic regions (Supplementary Figure 2c). Finally, most of H3K27ac peaks overlapped one or more ATAC-seq peaks (Supplementary Figure 2d). Together, we concluded 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 Figure 3), which induced thousands of peaks across all cell types (Figure 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 (Figure 1c); for example, in T cells, only 2% of H3K27ac and 0.8% of ATAC-seq peaks were condition specific. 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 (Figure 1c). The most variable peaks were in proximity to genes involved in cell differentiation, while low variability peaks were in proximity to genes involved in metabolic processes (Supplementary Figure 2e). We also observed that the levels of chromatin accessibility were less variable than those of H3K27ac. For instance, loci harboring hallmark genes for cytokine induced states showed marked differences in their H3K27ac profiles (Figure 1d) but not in ATAC-seq (Supplementary Figure 4). This is in line with other reports showing that chromatin activity marked by H3K27ac is more informative in discriminating between closely related cell states than chromatin accessibility[18,19]

Accounting for peak properties refines SNP enrichment

We next assessed if immune disease variants were enriched in chromatin elements of any specific cell state in our data. Typically, disease SNP enrichment analyses rely on the presence or absence of overlap between associated variants and regulatory regions[1,7,20]. However, in our data set 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 Figure 5). A similar observation was previously made using partitioning heritability analysis of neuropsychiatric and metabolic disorders in highly correlated chromatin annotations of different brain regions[5]. To assess the immune disease enrichment across the different T cell and macrophage states, we developed a new SNP enrichment method, *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 (Figure 2, Methods). Briefly, we first construct a matrix of quantile normalized read counts across peaks (Figure 2a). For each peak we generate specificity scores where 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; Figure 2b). To assess disease 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 (Figure 2c). Importantly, our method is peak-centric, i.e. 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 (Methods).

We used simulations to assess the sensitivity of *CHEERS* (Methods). With over 23% of variants overlapping the top 10% most cell-type-specific peaks, we observed over 80% power to detect a significant enrichment (p-value < 0.01). However, at least 78% SNPs 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 Figure 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. Based on the results from these simulations we concluded that *CHEERS* is able to identify enrichment in active chromatin elements across closely related cell types.

CHEERS identifies cell types relevant for immune diseases

To further validate *CHEERS,* we used a catalogue of 19 primary immune cell types assayed with H3K27ac ChIP-seq as a part of the BLUEPRINT project[3] and tested the enrichment of SNPs associated with 12 complex diseases with immune component, including: rheumatoid arthritis (RA), allergies, asthma, celiac disease (CED), inflammatory bowel disease (IBD), Crohn's disease (CD), ulcerative colitis (UC), multiple sclerosis (MS), psoriasis, type-1 diabetes (T1D), systemic lupus erythematosus (SLE) and systemic sclerosis. Concordant with previous results, we observed enrichment of rheumatoid arthritis variants in chromatin regions active in memory CD4+ T and CD8+ T cells[4,21] (Supplementary Figure 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 specifically active in memory CD4+ and CD8+ T cells. Moreover, we saw enrichment of risk variants for SLE and systemic sclerosis in chromatin elements specific to B cells (Supplementary Figure 7a) as previously reported[8].

We observed that the immune-mediated diseases that we tested fell into three groups (Supplementary Figure 7b): diseases with a strong T cell component (asthma, allergies, IBD, rheumatoid arthritis, CED, and T1D), diseases with a strong B cell component (psoriasis, SLE 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 SLE, *CHEERS* not only recapitulated previous enrichment in B cells but further refined the enrichment to class-switched B cells and not unswitched B cells (Supplementary Figure 7b). Based on these results, we concluded that *CHEERS* is also suitable for 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. We included three traits in this analysis: multiple sclerosis, coronary artery disease (CAD) and Parkinson's disease (Supplementary Figure 7d). We identified strong enrichment of GWAS loci for Parkinson's disease in cell types from different brain regions and the adrenal gland. Conversely, GWAS loci for multiple sclerosis showed significant enrichment in immune cell types. This showed that, *CHEERS* was sensitive to correctly disentangle the biological differences between multiple sclerosis and Parkinson's disease even though both affect the central nervous system. Moreover, we also identified a significant enrichment of risk variants for CAD in chromatin regions specific to the aorta (top enrichment), lungs and multiple muscle tissues (both skeletal and smooth). These results demonstrated that *CHEERS* was able to identify relevant cell type specific enrichments of GWAS loci for both immune and non-immune traits.

Refining cell states relevant for immune diseases

We next tested if GWAS variants associated with immune-mediated diseases were enriched in active chromatin regions specific to the cytokine polarizing cell states in our dataset. In addition, we included GWAS variants associated with Alzheimer's disease (AD), as microglia, a subtype of macrophages, were suggested to play a role in this disease[22,23]. The tested traits included from 17 to 132 associated loci, resulting in between 25 (asthma) and 317 (IBD) SNP-peak overlaps that contributed to the final enrichment scores.

Variants associated with multiple sclerosis, rheumatoid arthritis, CED, IBD, Crohn's disease, ulcerative colitis, psoriasis, T1D, allergy and asthma were predominantly enriched in T cells, which is in line with the observations from the BLUEPRINT data (Figure 3). This enrichment was particularly strong in chromatin regions specifically active during early activation of memory T cells (16 hours; Figure 3). We did not observe significant differences between early and late activation of T cells, either in the number or the quality (i.e. q-value) of H3K27ac peaks (Supplementary Figure 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 which drives these signals. We did not observe T cell enrichment in control traits such as coronary artery disease and low density lipoprotein (LDL) cholesterol (Figure 3). Finally, we observed that the enrichments were predominantly driven by peaks proximal to transcription start sites (TSS; Supplementary Figure 8).

The observed enrichment of immune variants in early memory T cell activation was driven by a group of H3K27ac peaks highly specific to this cell state (Figure 4a, Supplementary Figure 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 (Figure 4b). Allergy variants also mapped within a smaller group of peaks active specifically during early activation of naive T cells, explaining the secondary enrichment in activated naive T cells (Figure 4a). A proportion of immune disease variants also mapped in peaks specific to early activation of both naive and memory CD4 T cells. For instance, rheumatoid arthritis

associated variants overlapped a peak near acyl-CoA Oxidase-like protein (*ACOXL*) gene, which is involved in T cell metabolism (Figure 4b). Collectively, genes near peaks driving 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 (Figure 4d). This suggests that immune disease variants overlap enhancers and promoters that regulate gene expression programmes 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 variants were significantly enriched post IL-27 stimulation in early memory T cell activation ($p = 5.4 \times 10^{-10}$). This is concordant with previous studies that reported elevated levels of IL-27 in the cerebrospinal fluid of multiple sclerosis patients[24,25]. Among the late activation states, naive T cells polarized with Th1 cytokines particularly stood out as strongly significant for IBD, Crohn's disease, ulcerative colitis and CED ($p = 9.92 \times 10^{-9}$, $p = 1.11 \times 10^{-7}$, $p = 4.89 \times 10^{-4}$ and $p = 1.01 \times 10^{-5}$, respectively). The H3K27ac peaks intersecting risk variants for these diseases and with high specificity in Th1 cells were in close proximity to genes involved in Th1 biology, such as IFN-γ and IL-12R (Figure 4c). Also, collectively these genes were enriched in hallmark pathways of Th1 cells such as IFN-γ, IL-12 and JAK-STAT signaling (Figure 4d).

Across all tested traits, only Alzheimer's disease variants were significantly enriched in macrophages (Figure 3), however we did not observe any individual cytokine condition being more significant than others. Genes in proximity to peaks overlapping Alzheimer's disease associated variants driving the macrophage signal were enriched in pathways such as regulation of amyloid beta clearance and amyloid beta formation, processes extensively studied in the context of Alzheimer's disease pathology (Figure 4d).

Finally, we observed that in the ATAC-seq data the enrichment results were less significant (Supplementary Figure 10a) but the global patterns of the most relevant disease cell states replicated those from H3K27ac data (Supplementary Figure 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 TSS (Supplementary Figure 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 to 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 naive T cells with Th1 polarizing cytokines. Our approach provides a valuable framework for identifying disease relevant cell types and cell states and to inform the design of subsequent large-scale studies for functional interpretation of GWAS variants for complex immune disease.

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 relevant in the pathobiology of immune-mediated diseases. We observed that most changes in the chromatin landscape of CD4+ T cells resulted from TCR and CD28 stimulation alone, while 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 variants observed in previous studies[7,21,26] to specific cellular contexts. For ten of the twelve immune diseases that we tested, we observed that the associated variants were enriched in 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 identified enrichment of immune disease variants in T cell activation[27]. 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 identified elevated cytokine levels in patients with immune diseases and subsequently led to development of disease-modifying therapies targeting different cytokines[28–31]. This implies that effector T cells (corresponding to five-day stimulations in our dataset) and the cytokines that they secrete drive autoimmune inflammation. In 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, while cytokine induced cell states could be key in the later stages of pathologic inflammation, they might not be triggering a disease. This emphasises 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 where, 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[32–35]. 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. Therefore our results suggest that focusing on regulation of early activation of memory CD4+ T cell may provide an important axis for development of new treatment options.

In addition to early activation of memory CD4+ T cells, we observed selected diseases where 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 Th1 cytokines. This agrees with previous findings from immunophenotyping studies showing that lamina propria T cells express high levels of Stat4, T-bet and IL-12R, which are all induced by Th1 polarizing cytokines[36]. This could indicate that a proportion of IBD variants regulate the function of Th1 cells, suggesting the

involvement of this cell type in IBD development. Another example was asthma, where we detected enrichment in naive T cells polarized towards the Th2 phenotype.

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. In 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[22,23,37], a type of resident macrophage in the central nervous system (CNS). Our results suggest that a proportion of Alzheimer's disease associated variants could be studied in an *in vitro* model of monocyte differentiated macrophage. This can have significant implications, as cells from the CNS are challenging to collect. Finally, the lack of stratification of Alzheimer's disease variants between the different cytokine induced macrophage states could indicate that some Alzheimer's disease 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 chromatin regulatory landscape 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.

Methods

Sample processing and cell activation

Blood samples were obtained from nine healthy individuals. Cellular conditions within each cell type (naive 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 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 cell purity was above 94% in all samples (Supplementary Figure 1a). T cells were stimulated with anti-CD3/anti-CD28 human T-Activator Dynabeads® (Invitrogen) at 1 bead : 2 T cell ratio. Cytokines were added at the same time as the stimulus and cells were harvested after 16h and 5 days (for the list of cytokines, and their concentration refer to Supplementary Table 1).

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 (Supplementary Figure 1a). Following macrophage differentiation, they were stimulated with cytokines for 6 and 24 hours (for the list of cytokines, and their concentration refer to Supplementary Table 1).

Flow Cytometry

Cells were analyzed with BD LSR Fortessa. All data obtained by flow cytometry were analyzed by FlowJo Version 9.9.6 (TreeStar).

ChIPmentation-seq

In order 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 1×10^6 cells/ml in FACS buffer (PBS buffer supplemented with 10% FCS and 1 mM EDTA) and chromatin was cross-linked by adding 1% formaldehyde and incubating at 37°C for five minutes. To quench the reaction we added glycine and cells were washed 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. Briefly, cells were lysed and sonicated, and the resulting material was used for immunoprecipitation (IP). Sonication was performed using a Bioruptor® Pico (Diagenode), with 6 sound pulse cycles of 30 seconds each. For immunoprecipitation (IP), we used ChIP grade antibodies specific to human H3K27ac histones (Diagenode). A negative control undergoing no IP (input) was also generated for each cell type. After IP, the recovered chromatin fragments were tagmented as previously described[38]. Briefly, 2 µl of IP material were resuspended in 30 ul ChIP-seq buffer (Diagenode) supplemented with 1 µl Tn5 (Illumina). Samples were then reverse cross-linked using the iDeal® ChIP-seq kit for histones according to the manufacturer's instructions. The resulting DNA was purified using SPRI magnetic beads (AMPure XP A63881 Beckman Coulter). Enrichment of active chromatin regions was verified by qPCR.

Sequencing libraries were constructed from the obtained material using the Nextera® DNA library preparation kit according to the manufacturer's instructions. Briefly, DNA was amplified by PCR and fragments of inappropriate sizes were removed using Agencourt AMPure XP beads (BD). Finally, samples were pooled and loaded into an Illumina® HiSeq 2500 instrument for paired-end sequencing. In order to minimise batch effects, samples were randomized before sequencing. We obtained on average 63 million paired-end reads per sample.

ATAC-seq

In order 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 [39]. Briefly, 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 µl Tn5 transposase (Nextera). Samples were then incubated at 37°C and 800 rpm for 30 minutes. 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. Briefly, DNA was amplified by PCR and fragments of inappropriate sizes were removed using Agencourt AMPure XP beads (BD). Finally, samples were pooled and loaded into an Illumina® HiSeq 2500 for paired-end sequencing. In order to minimise batch effects, samples were randomized before library preparation and before sequencing. We obtained on average 58 million paired-end reads per sample.

ATAC-seq and ChM-seq analysis

We assessed the quality of reads using fastx and trimmed the adapters using skewer (v0.2.2)[40]. We then mapped reads to the human genome reference GRCh38 using bwa mem (v0.7.9a)[41] only keeping uniquely mapped reads. We also removed PCR duplicates and mitochondrial reads in ATAC-seq using samtools (v0.1.9)[41,42]. 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 (v2.6.0) to remove samples with over- or under- sonicated chromatin, and samples with skewed distributions of insert sizes.

Peaks were called using MACS2 (v2.1.1)[43]. 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*. For each cell type-stimulation pair we used an input from the same condition (for example, all memory 16H samples were called using input generated from memory 16H cells). We kept peaks that had q value < 10^{-2} in H3K27ac and q value < 10^{-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). The peak overlap between the assays was calculated using bedtools intersect (v2.22.0), while the distance of a peak to the closest transcription start site (TSS) and peak annotation was assessed using Homer's annotatePeaks.pl (v4.10.3).

To assess the quality of our data we calculated the fraction of reads in peaks (FRiP). We excluded 7 H3K27ac samples with FRiP less than 10% and 15 ATAC-seq samples with FRiP below 5%. The average q-values were 13.1 and 5.3 for ATAC-seq and ChM, respectively. The total number of samples remaining in our dataset after this QC was 173 in H3K27ac and 168 in ATAC. In addition samples with skewed distribution of insert sizes, or <10 million QCed read pairs 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 which 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[44]. We compared all conditions to the resting state or to Th0 and used Benjamini-Hochberg controlled false discovery rate (FDR) 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[3]. 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: allergies[45], Alzheimer's disease[46], asthma[47], celiac disease[48], coronary artery disease[49], levels of low density lipoprotein (LDL)[50], Crohn's disease[51], inflammatory bowel disease[51], multiple sclerosis[52], Parkinson's disease[53], psoriasis[54], rheumatoid arthritis[26], systemic lupus erythematosus[55], systemic sclerosis[56], type 1 diabetes[57] and ulcerative colitis[51]. Variants mapping to the HLA region were removed. SNPs were finally pruned by LD using a 50-kb window and $R^2 < 0.1$. Variants with minor allele frequency (MAF) lower 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[7]. We ran 10,000 permutations. Command used: *python goshifter.py -s SNP_list -a annotation/file/path -p 10000 -l LD/file/path -o output_name*.

Chromatin Element Enrichment Ranking Specificity (*CHEERS*)

QCed 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 the merged peak regions using featureCounts (v1.5.1)[58]. We normalized each peak for the library size by scaling the peak read counts to the sample with the highest count of informative sequence reads:

$$Z_{i,j} = C_{i,j} \frac{\max_i \left\{ \sum_j C_{i,j} \right\}}{\sum_j C_{i,j}}$$

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

To ensure that our analysis was not confounded by the low confidence peaks, we removed the bottom 10th percentile of peaks with the lowest read counts and obtained a final count of 127,723 peaks for H3K27ac ChM and 136,692 for ATAC-seq.

In order 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 that, 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:

$$S_{i,j} = \frac{Z_{i,j}}{\sqrt{\sum_i Z_{i,j}^2}}$$

where *Si, j* is the specificity score, and *Zi, j* is the normalized number of reads within a peak *j* in a condition *i*. This score is a number from 0 to 1, where 1 means 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, where the peak 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 the index variants, identify variants in LD (r2 > 0.8) and overlap them with peaks. Importantly, our method is peak-centric, i.e. 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 multiple peaks, each independent peak contributes to the final enrichment score. We then calculate the mean specificity rank of all peaks which 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 they follow a discrete uniform distribution (Kolmogorov-Smirnov test p=1) with mean (μ):

$$\mu = \frac{1+N}{2}$$

and variance:

$$\sigma^2 = \frac{(max - min + 1)^2 - 1}{12}$$

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:

$$\bar{\sigma}^2 = \frac{N^2 - 1}{12n}$$

where $\bar{\sigma}^2$ 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 $\Phi\left(\frac{x - \mu}{\sigma}\right)$ is a cumulative normal distribution for $x \sim N(\mu, \sigma^2)$; x is the observed rank, and μ and σ 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 that, within each of the tested cell states, we randomly sampled sets of peaks, matching for the number of peaks overlapping GWAS variants, and calculated the mean of their ranks. We repeated this process N times, assessed the frequency at which the mean of permuted ranks was greater or equal to the mean of observed ranks, and derived an empirical p-value:

$$p_{value} = \frac{num(permuted > observed)}{N}$$

where N is a number of permutations. Both approaches yielded similar results while the p-values from CHEERS were not limited by the number of permutations ($R^2 = 0.96$, p-value < 2.2×10^{-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-value < 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 10th percentile of specificity. Finally, in order 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).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was funded by the Open Targets (OTAR040). GT is supported by the Wellcome Trust (grant WT206194). ECG is supported by a Gates Cambridge Scholarship. LBC was supported by the MRC Skills Development Fellowship (MR/N014995/1).

References

1. Maurano MT, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012; 337:1190–1195. [PubMed: 22955828]
2. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. [PubMed: 22955616]
3. Chen L, et al. Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. Cell. 2016; 167:1398–1414.e24. [PubMed: 27863251]
4. Trynka G, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. Nat Genet. 2013; 45:124–130. [PubMed: 23263488]
5. Finucane HK, et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat Genet. 2018; 50:621–629. [PubMed: 29632380]
6. Fairfax BP, et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science. 2014; 343
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. 2015; 97:139–152. [PubMed: 26140449]
8. Farh KK-H, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2015; 518:337–343. [PubMed: 25363779]
9. Alasoo K, et al. Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. Nat Genet. 2018; 50:424–431. [PubMed: 29379200]
10. Martinez FO, Gordon S. The M1 and M2 paradigm of macrophage activation: time for reassessment. F1000Prime Rep. 2014; 6:13. [PubMed: 24669294]
11. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. Nature. 2011; 474:307–317. [PubMed: 21677747]
12. Yoshizaki A, et al. Regulatory B cells control T-cell autoimmunity through IL-21-dependent cognate interactions. Nature. 2012; 491:264–268. [PubMed: 23064231]
13. Mascanfroni ID, 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. 2013; 14:1054–1063. [PubMed: 23995234]
14. Feldmann M. Development of anti-TNF therapy for rheumatoid arthritis. Nat Rev Immunol. 2002; 2:364–371. [PubMed: 12033742]
15. Hall JC, Rosen A. Type I interferons: crucial participants in disease amplification in autoimmunity. Nat Rev Rheumatol. 2010; 6:40–49. [PubMed: 20046205]
16. Glinos DA, Soskic B, Jostins L, Sansom DM, Trynka G. Genomic profiling of T cell activation reveals dependency of memory T cells on CD28 costimulation. bioRxiv. 2018; doi: 10.1101/421099
17. Ye CJ, et al. Intersection of population variation and autoimmunity genetics in human T cell activation. Science. 2014; 345
18. Minoux M, et al. Gene bivalency at Polycomb domains regulates cranial neural crest positional identity. Science. 2017; 355
19. Mueller B, et al. Widespread changes in nucleosome accessibility without changes in nucleosome occupancy during a rapid transcriptional induction. Genes Dev. 2017; 31:451–462. [PubMed: 28356342]
20. Pickrell JK. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. Am J Hum Genet. 2014; 94:559–573. [PubMed: 24702953]

21. Hu X, et al. Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. Am J Hum Genet. 2011; 89:496–506. [PubMed: 21963258]
22. Calderon D, et al. Inferring Relevant Cell Types for Complex Traits by Using Single-Cell Gene Expression. Am J Hum Genet. 2017; 101:686–699. [PubMed: 29106824]
23. Tansey KE, Cameron D, Hill MJ. Genetic risk for Alzheimer's disease is concentrated in specific macrophage and microglial transcriptional networks. Genome Med. 2018; 10
24. Sénécal V, et al. Production of IL-27 in multiple sclerosis lesions by astrocytes and myeloid cells: Modulation of local immune responses. Glia. 2016; 64:553–569. [PubMed: 26649511]
25. Lalive PH, et al. Increased interleukin-27 cytokine expression in the central nervous system of multiple sclerosis patients. J Neuroinflammation. 2017; 14
26. Okada Y, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature. 2014; 506:376–381. [PubMed: 24390342]
27. Calderon D, et al. Landscape of stimulation-responsive chromatin across diverse human immune cells. bioRxiv. 2018; doi: 10.1101/409722
28. Schett G, Elewaut D, McInnes IB, Dayer J-M, Neurath MF. How cytokine networks fuel inflammation: Toward a cytokine-based disease taxonomy. Nat Med. 2013; 19:822–824. [PubMed: 23836224]
29. McInnes IB, Buckley CD, Isaacs JD. Cytokines in rheumatoid arthritis - shaping the immunological landscape. Nat Rev Rheumatol. 2016; 12:63–68. [PubMed: 26656659]
30. Babaloo Z, et al. Increased IL-17A but decreased IL-27 serum levels in patients with multiple sclerosis. Iran J Immunol. 2013; 10:47–54. [PubMed: 23502338]
31. Jamshidian A, Shaygannejad V, Pourazar A, Zarkesh-Esfahani S-H, Gharagozloo M. Biased Treg/Th17 balance away from regulatory toward inflammatory phenotype in relapsed multiple sclerosis and its correlation with severity of symptoms. J Neuroimmunol. 2013; 262:106–112. [PubMed: 23845464]
32. Kuehn HS, et al. Immune dysregulation in human subjects with heterozygous germline mutations in CTLA4. Science. 2014; 345:1623–1627. [PubMed: 25213377]
33. Lo B, et al. AUTOIMMUNE DISEASE. Patients with LRBA deficiency show CTLA4 loss and immune dysregulation responsive to abatacept therapy. Science. 2015; 349:436–440. [PubMed: 26206937]
34. Schubert D, et al. Autosomal dominant immune dysregulation syndrome in humans with CTLA4 mutations. Nat Med. 2014; 20:1410–1416. [PubMed: 25329329]
35. Tivol EA, 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. 1995; 3:541–547. [PubMed: 7584144]
36. Neurath MF, et al. The transcription factor T-bet regulates mucosal T cell activation in experimental colitis and Crohn's disease. J Exp Med. 2002; 195:1129–1143. [PubMed: 11994418]
37. Gosselin D, et al. An environment-dependent transcriptional network specifies human microglia identity. Science. 2017; 356
38. Schmidl C, Rendeiro AF, Sheffield NC, Bock C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat Methods. 2015; 12:963–965. [PubMed: 26280331]
39. Corces MR, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat Genet. 2016; 48:1193–1203. [PubMed: 27526324]
40. 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. 2014; 15
41. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics. 2010; 26:589–595. [PubMed: 20080505]
42. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. [PubMed: 19505943]
43. Zhang Y, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008; 9
44. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15

45. Ferreira MA, et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. Nat Genet. 2017; 49:1752–1757. [PubMed: 29083406]
46. Jansen IE, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat Genet. 2019; 51:404–413. [PubMed: 30617256]
47. Demenais F, et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. Nat Genet. 2018; 50:42–53. [PubMed: 29273806]
48. Trynka G, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat Genet. 2011; 43:1193–1201. [PubMed: 22057235]
49. Nelson CP, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. Nat Genet. 2017; 49:1385–1391. [PubMed: 28714975]
50. Hoffmann TJ, et al. A large electronic-health-record-based genome-wide study of serum lipids. Nat Genet. 2018; 50:401–413. [PubMed: 29507422]
51. Jostins L, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012; 491:119–124. [PubMed: 23128233]
52. International Multiple Sclerosis Genetics Consortium (IMSGC). et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nat Genet. 2013; 45:1353–1360. [PubMed: 24076602]
53. Chang D, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. Nat Genet. 2017; 49:1511–1516. [PubMed: 28892059]
54. Tsoi LC, et al. Large scale meta-analysis characterizes genetic architecture for common psoriasis associated variants. Nat Commun. 2017; 8
55. Langefeld CD, et al. Transancestral mapping and genetic load in systemic lupus erythematosus. Nat Commun. 2017; 8
56. Bossini-Castillo L, López-Isac E, Mayes MD, Martín J. Genetics of systemic sclerosis. Semin Immunopathol. 2015; 37:443–451. [PubMed: 26032405]
57. 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. 2015; 47:381–386. [PubMed: 25751624]
58. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014; 30:923–930. [PubMed: 24227677]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Editorial summary

Integration of immune disease GWAS variants with open chromatin and enhancer profiling in T cells and macrophages stimulated with different cytokines and analyzed at different time points identifies cell state-specific enrichments for 12 complex diseases.

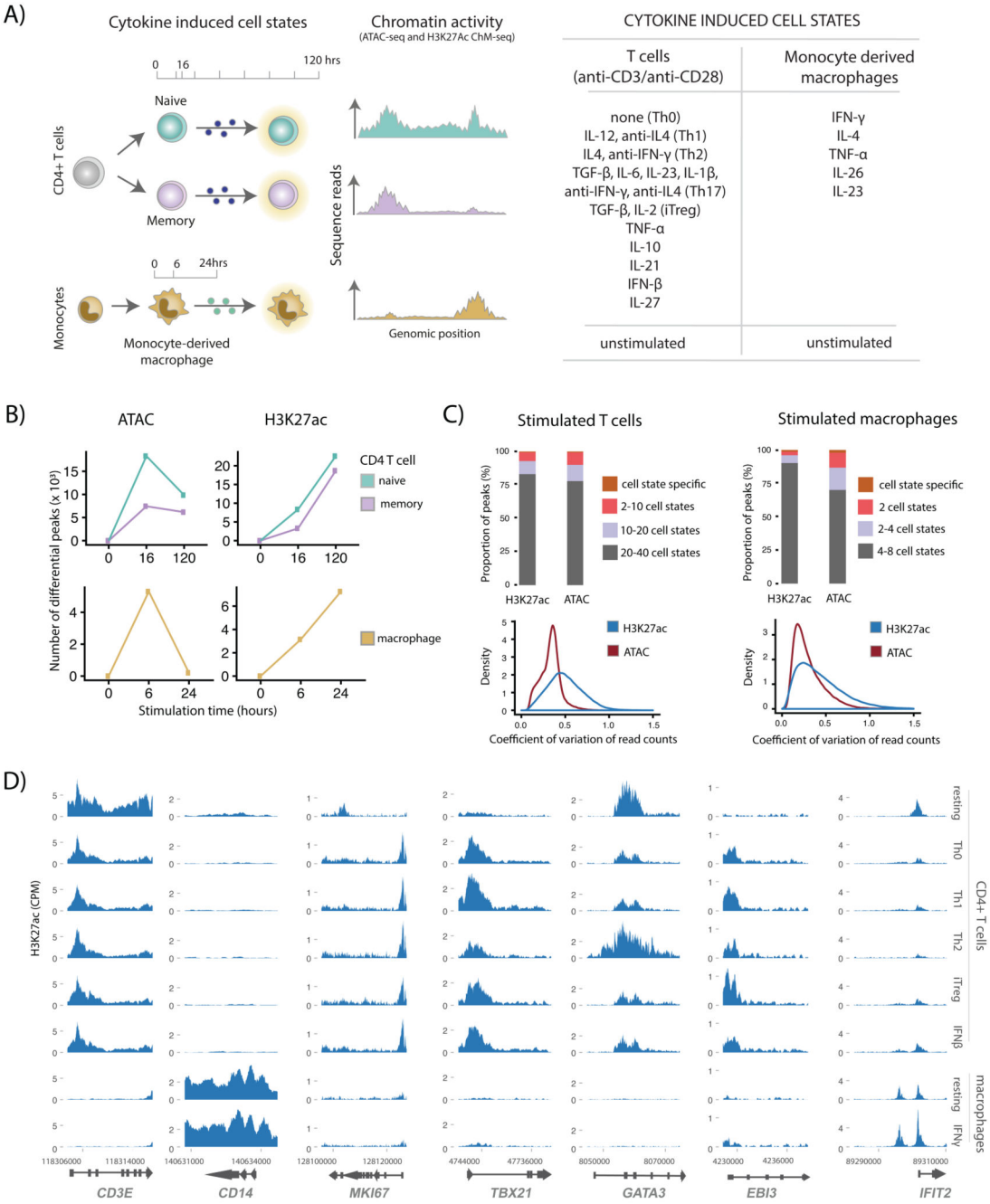

Figure 1. Quantitative changes in chromatin activity distinguish immune cell states.

a) Overview of the study design. Naive and memory CD4+ T cells were stimulated with anti-CD3/anti-CD28 T cell activation beads. Macrophages were differentiated from monocytes using M-CSF. 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 points. **b)** Number of differential H3K27ac and ATAC peaks upon T cell activation with anti-CD3/anti-CD28 T cell activation beads and macrophage activation with TNFa. **c)** Proportion of activation induced peaks that are shared within all T cell or macrophage states.

Europe PMC Funders Author Manuscripts

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), KI67 (T cell activation), TBX21 (Th1), GATA3 (Th2), IL35B (i.e. EBI3, iTreg), IFIT2 (IFN-induced). Genomic coordinates (GRCh38) for each gene are labeled. H3K27ac tracks are generated after merging three biologically independent samples per cell state.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

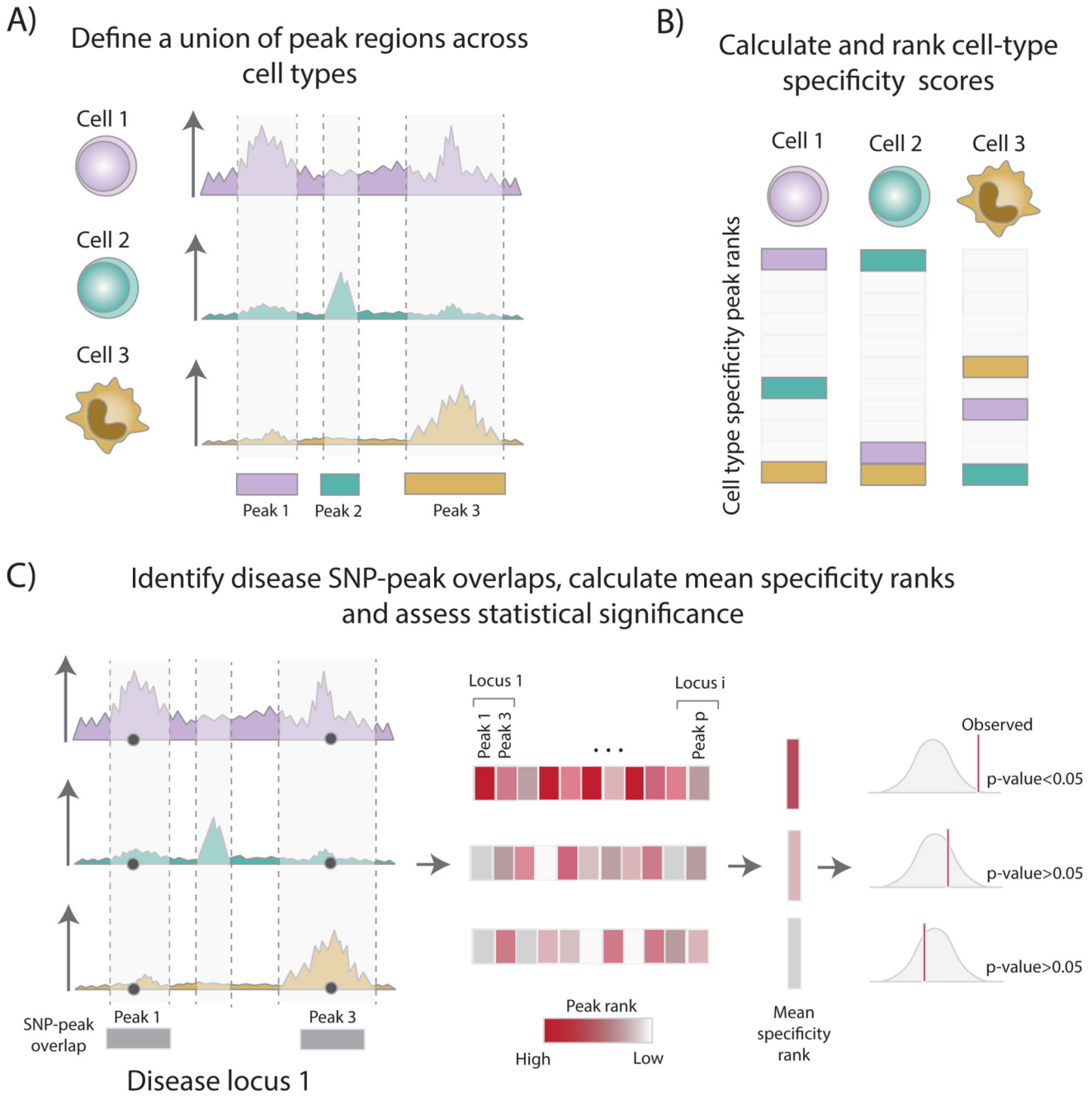

Figure 2. Overview of the *CHEERS* method.

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 disease SNP enrichment, we take all 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.

Europe PMC Funders Author Manuscripts

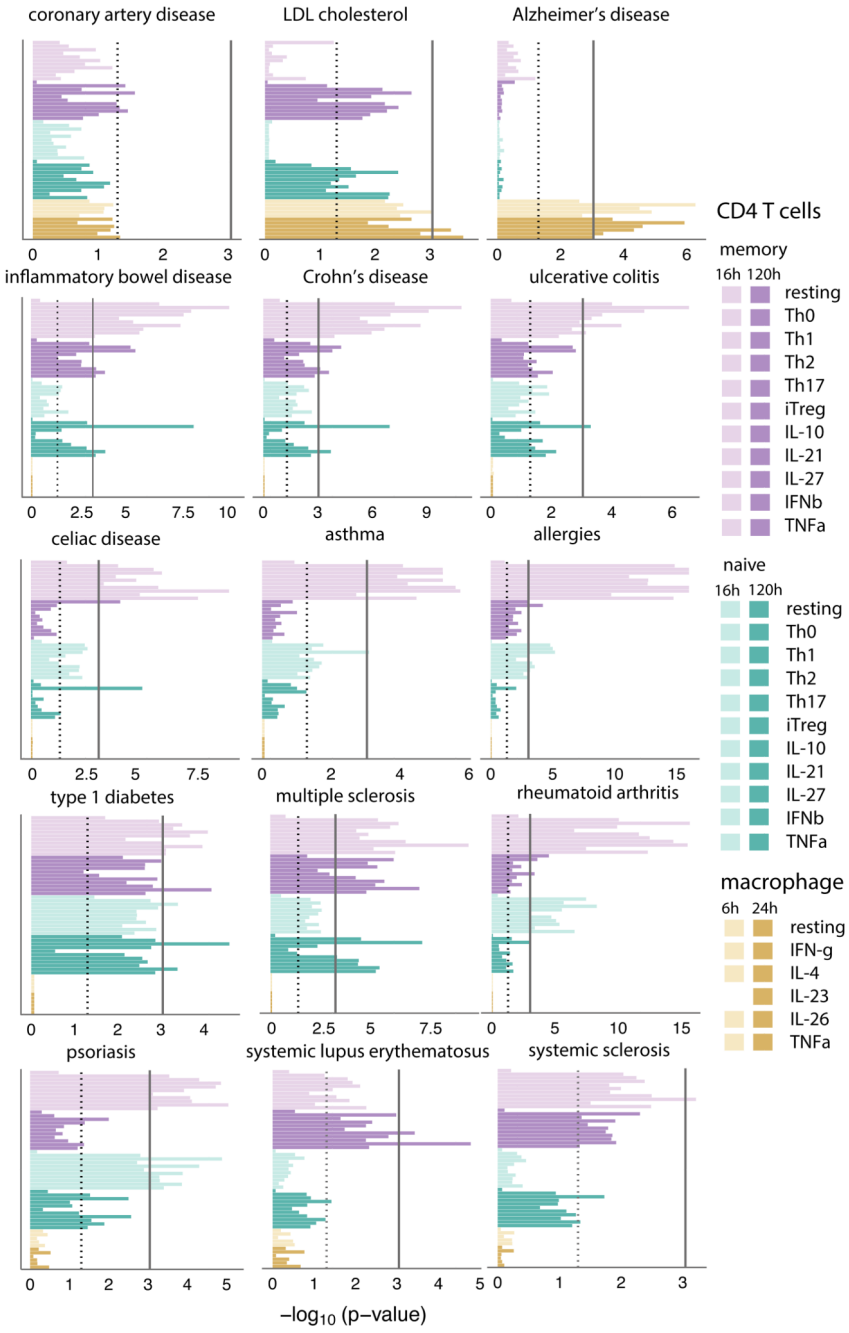

Figure 3. Disease SNP enrichment in H3K27ac regions in cytokine induced cell states.

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 "GWAS data processing" and "*CHEERS*" sections of the Methods. The dotted gray line marks the nominal p-value threshold of 0.05, while the solid gray line represents the Bonferroni-corrected significance threshold (p-value $\leq 9 \times 10^{-4}$).

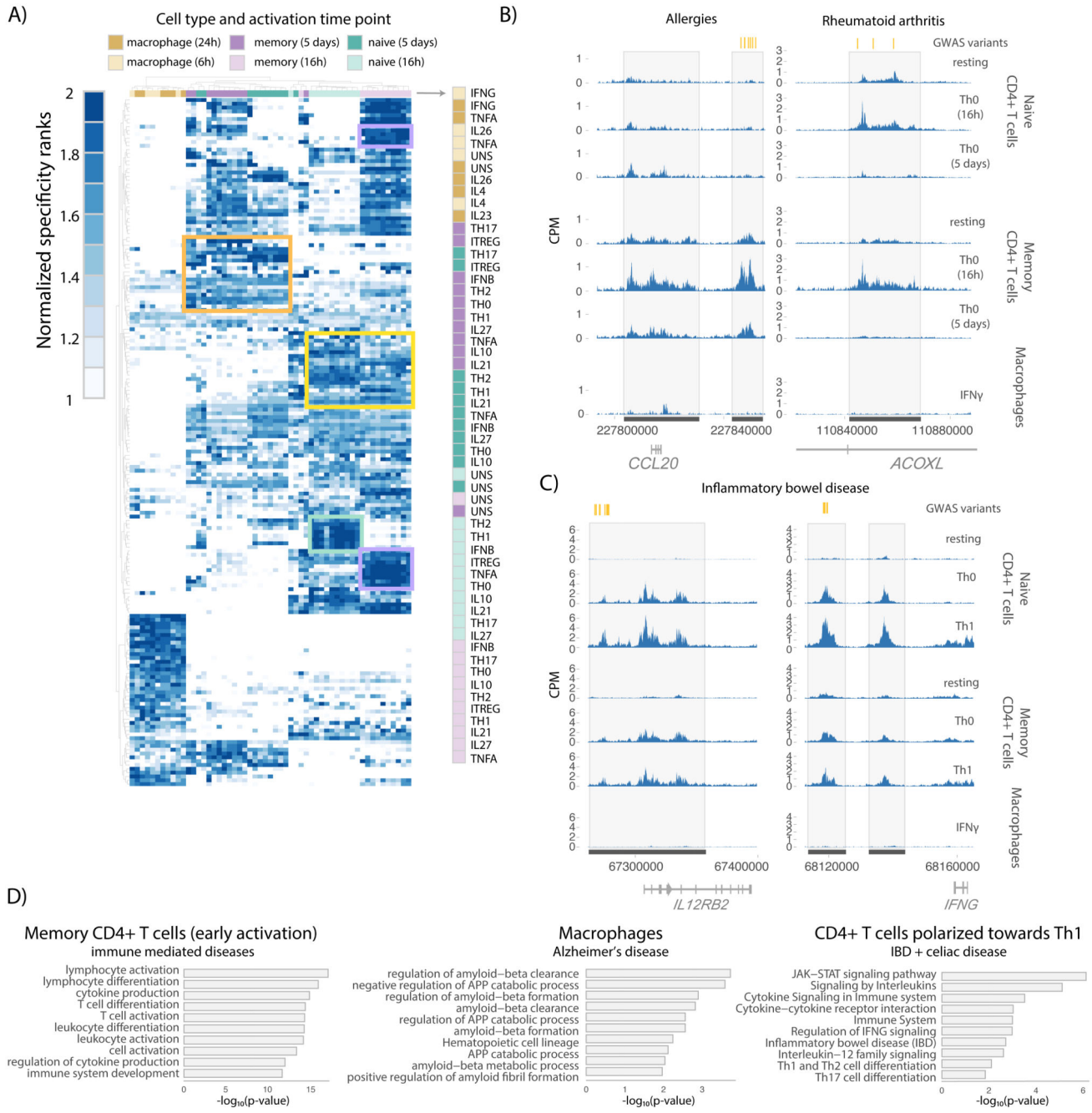

Figure 4. Example loci driving immune disease SNP enrichment in cytokine induced cell states.

a) Loci driving enrichment of allergy variants in early activation of memory T cells. Each row corresponds to a H3K27ac peak that intersects GWAS variants for allergy, while 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 pileups at selection of H3K27ac peaks driving the enrichment of allergy and rheumatoid arthritis variants in early T cell activation. Genomic coordinates (GRCh38)

Europe PMC Funders Author Manuscripts

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 variants in Th1-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), Th1 cells (IBD and CED) and macrophages (Alzheimer's disease). To determine these enrichments, we selected peaks with top 10% specific peaks at the enriched cell states and assigned them to the closest gene. P-values were calculated using g:ProfileR 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 Th1.