

RESEARCH

Open Access



The impact of cell type and context-dependent regulatory variants on human immune traits

Zepeng Mu^{1†}, Wei Wei^{2,3†}, Benjamin Fair⁴, Jinlin Miao^{2,3}, Ping Zhu^{2,3*} and Yang I. Li^{4,5*} 

*Correspondence:

zhuping@fmmu.edu.cn;

yangili1@uchicago.edu

[†]Zepeng Mu and Wei Wei contributed equally to this work.

²Department of Clinical Immunology, Xijing Hospital, Xi'an, China

⁴Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL, USA

Full list of author information is available at the end of the article

Abstract

Background: The vast majority of trait-associated variants identified using genome-wide association studies (GWAS) are noncoding, and therefore assumed to impact gene regulation. However, the majority of trait-associated loci are unexplained by regulatory quantitative trait loci (QTLs).

Results: We perform a comprehensive characterization of the putative mechanisms by which GWAS loci impact human immune traits. By harmonizing four major immune QTL studies, we identify 26,271 expression QTLs (eQTLs) and 23,121 splicing QTLs (sQTLs) spanning 18 immune cell types. Our colocalization analyses between QTLs and trait-associated loci from 72 GWAS reveals that genetic effects on RNA expression and splicing in immune cells colocalize with 40.4% of GWAS loci for immune-related traits, in many cases increasing the fraction of colocalized loci by two fold compared to previous studies. Notably, we find that the largest contributors of this increase are splicing QTLs, which colocalize on average with 14% of all GWAS loci that do not colocalize with eQTLs. By contrast, we find that cell type-specific eQTLs, and eQTLs with small effect sizes contribute very few new colocalizations. To investigate the 60% of GWAS loci that remain unexplained, we collect H3K27ac CUT&Tag data from rheumatoid arthritis and healthy controls, and find large-scale differences between immune cells from the different disease contexts, including at regions overlapping unexplained GWAS loci.

Conclusion: Altogether, our work supports RNA splicing as an important mediator of genetic effects on immune traits, and suggests that we must expand our study of regulatory processes in disease contexts to improve functional interpretation of as yet unexplained GWAS loci.

Keywords: eQTL, sQTL, GWAS, Colocalization



Background

Genome-wide association studies (GWAS) have identified well over ten thousand genomic loci associated with human diseases and complex traits [1]. However, while the number of trait-associated variants continues to grow, the causal genes and mechanisms at most GWAS loci remain difficult to determine. This difficulty is in part owing to the fact that ~90% of GWAS variants lie in noncoding regions [2].

Multiple studies have now shown that noncoding trait-associated variants are enriched for expression QTLs (eQTLs) and enriched in regulatory elements such as enhancers and promoters [3–5]. These findings suggest that noncoding variants likely affect traits by impacting gene regulation, an interpretation which has motivated many studies to map regulatory QTLs—in particular eQTLs—in a diverse set of tissues and cell types [6–13]. While eQTLs indeed overlap with many variants that have been associated with complex traits and diseases [3], several studies that assessed colocalization between GWAS and eQTL variants concluded that only a minority of GWAS loci can be explained by the eQTLs detected in available samples. For example, a 2017 study [14] reported that ~21% of variants associated with autoimmune diseases colocalize with eQTLs in at least one of three immune cell types they analyzed. In addition, a paper from the GTEx consortium [15] suggests that ~20% of GWAS loci show colocalized effect with eQTLs in the tissue most relevant to the trait. Moreover, another recent study estimated that only an average of 11% of trait narrow-sense heritability could be explained by *cis*-genetic effects on gene expression levels as measured in GTEx [16]. Altogether, these observations suggest that very little is known about the genes and mechanisms by which genetic variants impact traits at the vast majority of GWAS loci.

There are several possible explanations for the modest overlap between GWAS loci and eQTLs. For example, there may exist genetic effects on gene regulatory processes other than steady state gene expression levels that mediate genetic effects on trait. Indeed, we previously showed that RNA splicing is an important regulatory mechanisms that link trait-associated variants to complex traits [17]. Another explanation is that genetic effects are often restricted to trait-relevant cell types and cell states that have not been the subjects of colocalization or eQTL studies. Indeed, because the effect of GWAS loci on gene regulation can be cell type-specific, QTL maps in precise trait-relevant cell types must be available for successful colocalization. Additionally, the effects of GWAS loci have also been reported to be disease-specific [18], and can be found only when QTL mapping in samples collected from disease patients is available. Finally, genetic effects on gene regulation may sometimes be too small to be detected at current sample sizes, even in the causal cell types, cell states, and disease context. While all these possibilities likely contribute to the modest overlap that has been observed, identifying major contributors would significantly help our design of future human genomics studies.

The large number of GWAS loci without a colocalized eQTL is particularly striking for immune-related trait GWAS given that immune cell types have been the subject of the most eQTL studies, and with the largest sample sizes for eQTL mapping. This study aims to leverage the large number of eQTL studies available for immune cell types to understand how regulatory variants affect common disease risk, with a particular focus on the ~80% of autoimmune disease GWAS loci without a colocalized eQTL. Our approach was to perform a uniform eQTL and splicing QTL (sQTL) mapping across cell types and

datasets to evaluate the prevalence of cell type and cell state-specific effects, as well as to quantify the colocalization rates between sQTLs and GWAS loci.

To this end, we mapped eQTLs and sQTLs in four datasets, including (i) a dataset with a large number of different immune cell types but a small sample size (DICE [8], $N = 90$), (ii) a dataset with a single tissue-type but with a large sample size (DGN, $N = 922$ [10]), and (iii) two intermediate datasets (BLUEPRINT, $N = 197$ [7], and GEUVADIS, $N = 462$ [9]). We reasoned that analyzing these datasets in a uniform fashion (Fig. 1a) would allow us to capture both strong QTLs with cell type-specificity (using DICE) and weak-effect eQTLs that are less likely to be cell type-specific (e.g., using DGN). These data allowed us to evaluate the cell type-specificity of QTLs while considering limited statistical power due to small sample size [19, 20]. Moreover, using the uniformly processed QTLs, we found that, on average, eQTLs and sQTLs together colocalize with 40.4% of GWAS loci for autoimmune diseases and blood-related phenotypes, doubling for many GWAS the number of colocalizing loci from previous studies [14]. Interestingly, we found that cell type-specific eQTLs account for a very small number of colocalization events, a finding that stands in contrast to several previous studies [7, 8]. Notably, we found that genetic effects on RNA splicing contributed a large number of novel colocalizations, implying that RNA splicing is often impacted by trait-associated variants.

To characterize the remaining 60% of GWAS loci without colocalization, we collected H3K27ac profiles of 5 immune cell types (CD4⁺ T cells, CD8⁺ T cells, regulatory T cells, monocytes, and B cells) from rheumatoid arthritis (RA) patients and healthy controls using CUT&Tag [21]. These additional data helped us to better understand the cellular context in which GWAS loci, including those without colocalization, contribute to disease risk. Specifically, our work suggests that to understand the mechanisms underlying many GWAS risk loci, we need to study gene regulation in the context of disease. This stands in contrast to the idea that gene regulation in the disease context merely reflects the consequence of disease, or of response to treatment, which would confound the study of causal genetic mechanisms rather than help it.

Taken together, our work reports a comprehensive analysis of the regulatory effects of genetic variants on immune cell types, their overlap with GWAS loci and with regulatory regions in a disease context. Our maps of eQTLs, sQTLs, and gene regulatory regions in diverse immune cell types are available online [22], which we foresee will aid research on the genetic basis of diseases and on gene regulation in immune cells.

Results

Harmonized map of eQTLs and sQTLs in 18 immune cell types

We built a uniform data processing pipeline to harmonize four population-scale RNA-seq datasets (Fig. 1a). The DICE dataset consists of population RNA-seq data for 13 unstimulated immune cell types including various naïve and effector/memory T cell subtypes, classical and non-classical monocytes, B cells, and NK cells. The DICE dataset also includes RNA-seq data from CD4⁺ and CD8⁺ T cells that have been activated in vitro by engaging T cell receptor (TCR) complex using CD3/CD28 antibodies. Although the sample size in the DICE dataset is the smallest ($n = 91$) among the four datasets, the large number of sorted cell types makes the DICE dataset ideal to identify cell type-specific genetic effects. The BLUEPRINT dataset consists of RNA-seq data from three cell types (classical monocytes, naïve CD4⁺ T cells and neutrophils) in ~197 individuals. The DGN

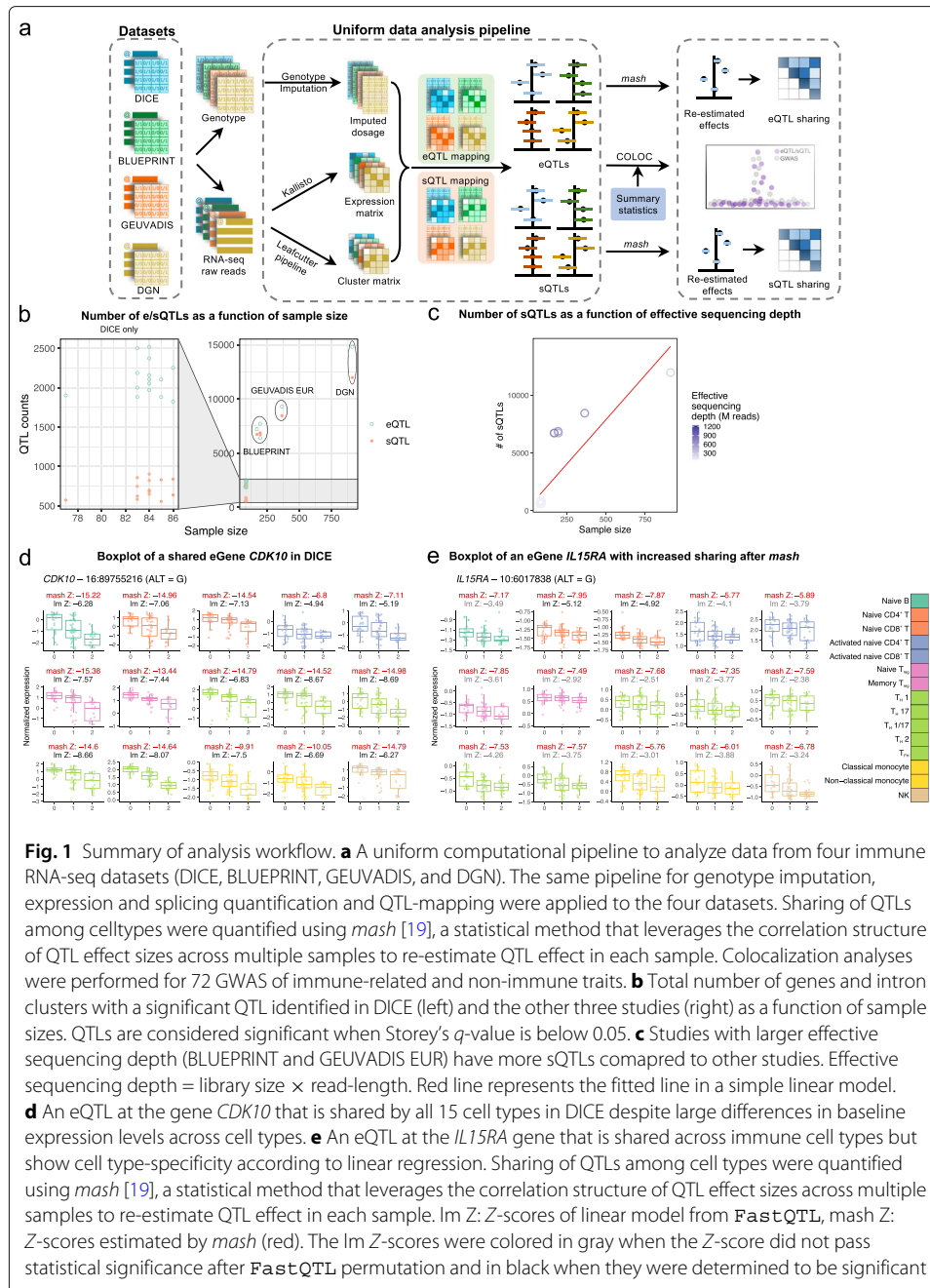


Fig. 1 Summary of analysis workflow. **a** A uniform computational pipeline to analyze data from four immune RNA-seq datasets (DICE, BLUEPRINT, GEUVADIS, and DGN). The same pipeline for genotype imputation, expression and splicing quantification and QTL-mapping were applied to the four datasets. Sharing of QTLs among celltypes were quantified using *mash* [19], a statistical method that leverages the correlation structure of QTL effect sizes across multiple samples to re-estimate QTL effect in each sample. Colocalization analyses were performed for 72 GWAS of immune-related and non-immune traits. **b** Total number of genes and intron clusters with a significant QTL identified in DICE (left) and the other three studies (right) as a function of sample sizes. QTLs are considered significant when Storey’s *q*-value is below 0.05. **c** Studies with larger effective sequencing depth (BLUEPRINT and GEUVADIS EUR) have more sQTLs compared to other studies. Effective sequencing depth = library size × read-length. Red line represents the fitted line in a simple linear model. **d** An eQTL at the gene *CDK10* that is shared by all 15 cell types in DICE despite large differences in baseline expression levels across cell types. **e** An eQTL at the *IL15RA* gene that is shared across immune cell types but show cell type-specificity according to linear regression. Sharing of QTLs among cell types were quantified using *mash* [19], a statistical method that leverages the correlation structure of QTL effect sizes across multiple samples to re-estimate QTL effect in each sample. Im Z: Z-scores of linear model from *FastQTL*, mash Z: Z-scores estimated by *mash* (red). The Im Z-scores were colored in gray when the Z-score did not pass statistical significance after *FastQTL* permutation and in black when they were determined to be significant

consortium collected whole blood samples from 922 individuals and, finally, GEUVADIS collected RNA-seq data from 462 lymphoblastoid cell lines (LCL).

Our pipeline, which is described in detail in the “[Methods](#)” section, includes quantifying RNA expression and splicing levels, imputing genotype data to the same common reference panel, and calling QTLs in all datasets using the same strategy. We also designed an approach to harmonize quantification for splicing junction usage across cell types and datasets by first merging *LeafCutter* intron clusters [23] across all samples and then re-calculating intron usage for each sample (“[Methods](#)” section). Thus, we produced a harmonized set of introns that can be readily interrogated.

To map eQTLs and sQTLs, we used *FastQTL* [24]. As covariates for the linear regression, we used three genotype PCs and a number of phenotypic PCs chosen to maximize the number of significant QTLs (Storey's q -value < 0.05) (Additional File 1: Table S1-2). In total, we discovered 26,271 genes and 23,121 intron clusters that have a significant QTL in at least one of the four datasets at 5% false discovery rate (FDR). As expected, both the numbers of eQTLs and sQTLs were correlated with sample size (Fig. 1b). In addition to sample size, we found that the number of sQTLs identified was also correlated with effective sequencing depths (Fig. 1c and Figure S1). Notably, while the number of sQTLs is roughly linearly related to sample size, datasets with higher effective sequencing depths consistently yielded more sQTLs than predicted by a simple linear model. This is most obvious for BLUEPRINT, which used 100 bp single-end or paired-end sequencing when compared to DICE or DGN (both 50 bp single-end).

We show a shared eQTL for the *CDK10* gene (Fig. 1d) and an eQTL for the *IL15RA* gene (Fig. 1e) as examples. All gene expression and splicing quantifications, as well as all identified eQTLs and sQTLs are available on Zenodo [22] (see Additional File 1: Table S8).

Global patterns of eQTL and sQTL sharing across immune cell types

To characterize the cell type-specificity of genetic effects on gene regulation in immune cells, we sought to discern genetic variants that impact gene regulation broadly across many or all immune cell types from those that impact a few or only one cell type. Previous studies have also quantified the sharing and specificity of regulatory QTLs [7, 8]. However, because the sample sizes of most datasets are small, we speculated that estimates of QTL effect sizes are noisy, which would generally cause studies to underestimate the levels of QTL sharing.

We reasoned that our harmonized dataset would allow us to better infer sharing patterns. In particular, we improved our estimates of eQTL and sQTL effect sizes at each locus by statistical shrinkage using *mash* [19]. The *mash* method improves estimates of QTL effect sizes from those that are obtained from applying linear regression in each cell type separately, because *mash* leverages the correlation structure of QTL effect sizes across all cell types to re-estimate QTL effect sizes at each locus. We applied *mash* to calculate posterior mean effect sizes (henceforth referred to as *mash* effect sizes) and corresponding standard errors for the 36,950 unique SNP-gene associations and 116,881 unique SNP-intron associations (q -value below 5%) in the 15 DICE cell types separately. This procedure greatly enhanced estimates of QTL effect sizes in the 15 immune cell types (for two examples see Z -scores in Fig. 1d, e, also see Figure S2 and “Methods” section).

We first asked about the proportion of QTLs that are shared across immune cell types based on the estimated *mash* effect sizes. We found that a large fraction (33.7%, $n = 2897$ of 8597) of genes with an eQTL (eGenes) are shared according to *mash* (Local False Sign Rate [LFSR] < 0.05) across all six distinct major cell types in the DICE dataset (B cell, naïve CD4⁺ and CD8⁺ T cell, NK cell, classical monocytes, non-classical monocytes) (Fig. 2a). Our estimates of sharing are therefore much higher than the 5.2% (463 out of 8863) estimated in the original DICE study [8]. In fact, the original DICE study estimated that nearly half of all eGenes are specific to a single immune cell type, while our new estimate suggests that only 20.4% are likely cell type-specific (Fig. 2a).

Using *mash* effect sizes, it is also possible to quantify the amount of QTL sharing in terms of magnitude of effects. We found that over 40% of eQTLs have similar *mash*

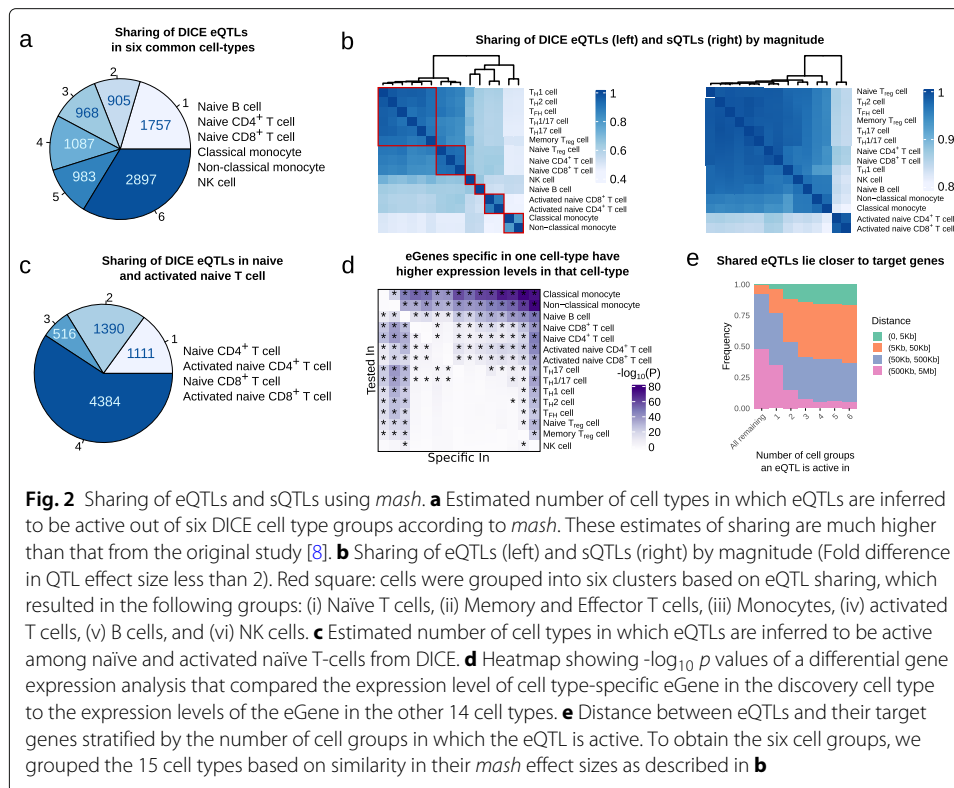


Fig. 2 Sharing of eQTLs and sQTLs using *mash*. **a** Estimated number of cell types in which eQTLs are inferred to be active out of six DICE cell type groups according to *mash*. These estimates of sharing are much higher than that from the original study [8]. **b** Sharing of eQTLs (left) and sQTLs (right) by magnitude (Fold difference in QTL effect size less than 2). Red square: cells were grouped into six clusters based on eQTL sharing, which resulted in the following groups: (i) Naive T cells, (ii) Memory and Effector T cells, (iii) Monocytes, (iv) activated T cells, (v) B cells, and (vi) NK cells. **c** Estimated number of cell types in which eQTLs are inferred to be active among naïve and activated naïve T-cells from DICE. **d** Heatmap showing $-\log_{10} p$ values of a differential gene expression analysis that compared the expression level of cell type-specific eGene in the discovery cell type to the expression levels of the eGene in the other 14 cell types. **e** Distance between eQTLs and their target genes stratified by the number of cell groups in which the eQTL is active. To obtain the six cell groups, we grouped the 15 cell types based on similarity in their *mash* effect sizes as described in **b**

effect sizes (within 2-fold, 34% within 1.5-fold) across pairs of cell types, and this fraction increases to over 90% when considering closely related cell types, such as classical and non-classical monocytes or T_H1 and T_H17 cells (Fig. 2b, left). In addition, we found that the vast majority (> 80% within 2-fold, 67% within 1.5-fold) of sQTLs have similar *mash* effect sizes across all immune cell types, with activated CD4⁺ and CD8⁺ T cells forming an outlier group in a hierarchical clustering based on estimates of sharing (Fig. 2b, right). These results are consistent with previous work [6] and suggest that the impact of genetic variation on RNA splicing is generally shared across two cell types when the involved mRNA transcripts are expressed in both cell types, which is largely the case for any pair of immune cell types.

In general, the proportion of shared eQTLs across cell types captured the lineage relationships among the 15 immune cell types. Specifically, classical and non-classical monocytes clustered together, while B cells and NK cells each formed distinct clusters. Furthermore, despite a high level of QTL sharing (> 80%) among naïve T cells, we found that naïve CD4⁺, CD8⁺ and regulatory T cells formed one cluster, while memory and effector T cells formed another larger cluster. We also observed a higher level of QTL sharing between activated CD4⁺ and CD8⁺ T cells compared to that between stimulated and naïve T cells. This observation suggests that activated CD4⁺ and CD8⁺ T cells share similar gene expression programs upon activation, and that differences in genetic effects on gene regulation exist between activated and non-activated cells. Nevertheless, we found that 66.2% of eGenes ($n = 4900$) were shared according to *mash* (LFSR < 0.05) between naïve and activated T cells, suggesting that the overall impact of genetic effects on gene regulation is in most cases the same across activated and naïve T cells (Fig. 2c,

Additional File 1: Table S3). We also calculated share-by-sign and share-by-magnitude excluding the HLA locus (Chr6: 25–35 Mb), and observed no significant difference in eQTL sharing levels. This is not surprising given that the HLA locus contains but a few hundreds genes, which account for 1.26% of all genes included in our *mash* analysis (Figure S3).

While a large proportion of eQTLs and sQTLs appeared to be shared across multiple immune cell types, we found that a substantial number of eQTLs (2810 eQTLs, 27.8%, which include stimulation-specific eQTLs) appeared cell type-specific. We asked whether QTLs that appeared cell type-specific showed specific features compared to QTLs that were shared across immune cell types. We first asked whether genes with eQTLs that were specific to a cell type were also more highly expressed in that cell type compared to the other cell types. To test this, we asked whether genes with an eQTL in a cell type *A* but not in another cell type *B*, were significantly more highly expressed in cell type *A* than cell type *B*. Indeed, we found that this was the case for most cell type-specific eQTLs (66.7%, Bonferroni adjusted *P* value <0.05, one-sided, paired Wilcoxon rank-sum test), suggesting that variation in gene expression level likely impacts whether a genetic variant has a regulatory effect and/or our ability to detect this effect. This observation was most obvious for classical monocytes, non-classical monocytes and naïve B cells, and is driven by differences in their gene expression levels compared to T cells (Fig. 2d). In addition to differences in gene expression levels, we found that eQTLs that were cell type-specific were located further away from the gene transcription start site in comparison to eQTLs that were shared across immune cell types (Fig. 2e). Moreover, cell type-specific eQTLs were more highly enriched in enhancers compared to eQTLs that were shared (Figure S4). These observations are consistent with the notion that cell type-specific eQTLs tend to impact enhancer activity, while shared eQTLs more often impact promoters [25].

Taken together, our analyses revealed that QTL effects are shared for a large number genes. Nevertheless, we were able to detect a non-negligible number of cell type or cell group-specific QTLs. Importantly, these findings and classification show replication across datasets (Additional File 1: Supplementary Note 1). Thus, we expect our QTL data to be highly replicable in existing or future immune QTL datasets.

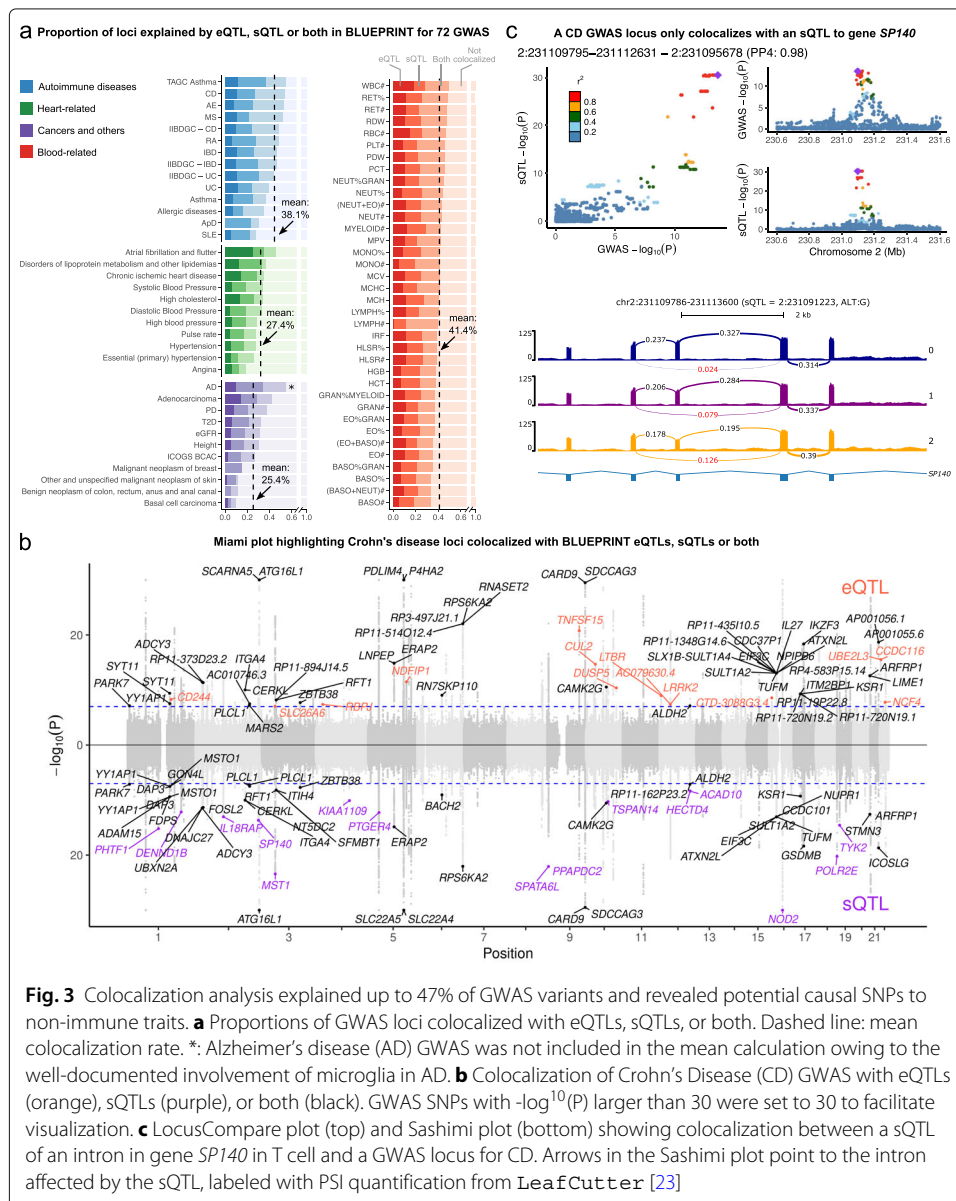
Colocalization of immune regulatory QTLs with common disease GWAS

Our harmonized eQTL and sQTL data gave us the unprecedented ability to identify genetic variants that impact traits through regulatory effects on immune cell types. We performed colocalization analyses that aimed to determine whether the genetic variants at GWAS loci that are causal for a trait are likely to be the same variants as the causal regulatory QTLs. We compiled a set of 72 well-powered GWAS, including 14 for autoimmune diseases (11 unique disease types), 36 blood traits, and 22 other traits (Additional File 1: Table S4), and used COLOC to evaluate colocalization ($PP4 \geq 0.75$) [26] with DICE, BLUEPRINT, and DGN QTLs separately (Additional File 3; average $N = 206,090$). We computed the colocalization rate for each GWAS as the percentage of GWAS loci that show evidence of colocalization out of the total number of associated loci in the GWAS (“Methods” section). We report the main colocalization results of our analyses using BLUEPRINT QTLs (3 immune cell types) below, and use the DICE (15 cell types) regulatory QTLs to interpret the cell type-specificity of colocalized genes. We reasoned that choosing BLUEPRINT over DICE as the main dataset for this anal-

ysis will increase our power for QTL mapping owing to its larger sample size and will also allow us to identify more sQTLs owing to higher RNA-seq coverage and longer read-lengths (Fig. 1c).

When we ascertained colocalization between GWAS loci for the 72 traits and QTLs from BLUEPRINT, we observed that colocalization rates between immune regulatory QTLs and GWAS hits were higher for autoimmune and blood-related traits compared to other non-immune traits (mean 40.4% versus 27.7%) (Fig. 3a). This observation supports the expectation that a large fraction of colocalized regulatory QTLs indeed affect immune traits by impacting gene regulation in immune cell types.

We next focused on autoimmune diseases and blood-related traits. Our regulatory QTLs colocalized with a mean of 38.1% (range: 24–47.4%, $n = 14$) and 41.4% (range: 33.8–50%, $n = 36$) of autoimmune disease and blood traits GWAS loci ($PP4 > 0.75$), respectively



(Fig. 3a). The mean rates of colocalization ranged from 27.4% to 50.2% depending on the choice of posterior probability cutoff for determining colocalization status (PP4, ranging from 0.5 to 0.9, Figure S5). We chose to use an intermediate cutoff of 0.75 to be consistent with previous studies [14]. Expression QTLs colocalized with 6.7–39.3% of GWAS loci with an average of 26.6%, similar to estimates from a previous study [14]. Notably, we found that splicing QTLs colocalized with an additional 7.6–21% GWAS loci (average: 13.8%) that did not colocalize with an eQTL, and explain much of the increase in colocalization rates from this study compared to that of previous studies. Interestingly, we observed that the rates of colocalization between GWAS loci and both an eQTL and an sQTL can vary substantially across traits, ranging from 4.5% for systemic lupus erythematosus (SLE) to 28.6% for basophil percentages of granulocyte (BASO%GRAN) (average: 17.2%). Most notably, nearly all colocalized loci associated with SLE (10 out of 16) colocalized only with sQTL (Fig. 3a). Interestingly, the rates of colocalization were not correlated with GWAS sample size nor the number of significant loci, and thus the variation in colocalization rates cannot be attributed to differences in GWAS power (Figure S6). This result raises the possibility of distinct regulatory architectures for different diseases. We obtained similar rates of colocalization with DICE and DGN, for which 30.7% and 38% immune GWAS loci colocalized with DICE and DGN regulatory QTLs, respectively (Figure S7).

To help the interpretation of these results, we show the colocalizations between immune regulatory QTLs and GWAS loci for Crohn's disease (CD) as an example [27] (Fig. 3b) (12,194 cases and 28,072 controls). We included 108 GWAS loci in our colocalization analysis that pass a p -value threshold of 10^{-7} (see "Methods" section). Ten and fifteen loci colocalized with only eQTLs or sQTLs, respectively, while an additional 25 loci colocalized with both eQTLs and sQTLs. In total, 46% of loci colocalized with an eQTL, an sQTL, or both. Of note, several identified colocalized genes have been extensively studied in terms of CD etiology, including *NOD2* [28] and *ITGA4*, of which the latter is the target for the CD monoclonal antibody drug natalizumab [29].

The high rates of colocalization (average: 13.8%) between GWAS loci and sQTLs highlight the importance of considering the impact of risk variants on RNA splicing. For example, we identified an sQTL associated with the skipping of the seventh exon in gene *SP140* in T cells that colocalized with a risk locus in both CD GWAS we analyzed (Fig. 3c) [27, 30]. *SP140* encodes nuclear body protein SP140 [31], which preferentially binds to gene promoters with H3K27me3 modification [32] and regulates multiple immune-related genes [33]. Notably, the exclusion of the same exon in *SP140* transcript isoforms has also been associated with risk alleles for other diseases including multiple sclerosis [34].

As expected, immune regulatory QTLs colocalized at a lower rates in GWAS of traits that are not autoimmune or blood-related (27.7%). Among the 22 non-immune traits we analyzed, Alzheimer's disease (AD) is an outlier, for which 55% of GWAS loci colocalized with a BLUEPRINT QTL. The high rate of colocalization can be explained by the known role of microglia in AD etiology [35]. Nevertheless, it is likely that for most other non-immune traits GWAS loci, colocalization with immune regulatory QTLs reflect a causal effect of the risk variant on disease through non-immune cell types that is also manifested in an immune cell type (see Additional File 1: Supplementary Note 2).

GWAS-eQTL colocalizations across immune cells are highly shared when accounting for statistical power

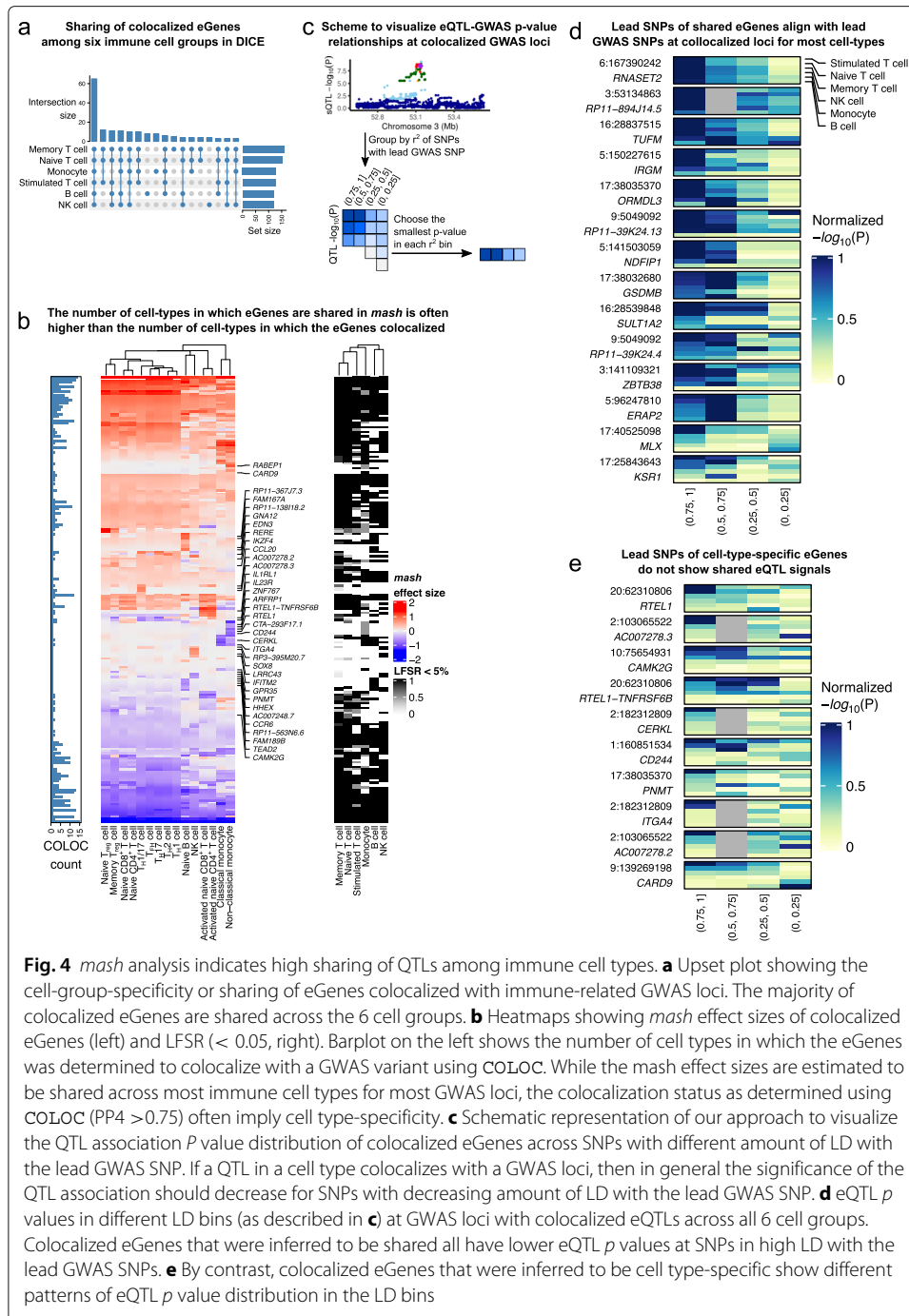
Several studies have proposed that a large fraction of autoimmune disease risk loci affect gene expression levels in a cell type-specific manner [36, 37]. We sought to use our dataset to evaluate this hypothesis by analyzing the cell type-specificity of the eQTLs that colocalize with autoimmune GWAS loci. To do this, we focused on the 197 genes with a DICE eQTL (eGenes) that colocalized with at least one of the 14 autoimmune disease GWAS in our study. We then evaluated the cell type-specificity using the *mash* QTL effect sizes estimated for the 15 immune cell types from the DICE consortium.

The general pattern of sharing that we observed for colocalized risk loci is that the corresponding eGenes are mostly shared across multiple cell types. The sharing was also apparent across the 6 major groups of immune cells that represent naïve T cells, memory and effector T cells, monocytes, activated T cells, B cells, and natural killer (NK) cells (Fig. 4a). Sixty five of 197 (33.0%) tested genes colocalized in all 6 major immune cell groups. The immune cell groups in which the most colocalized genes were found are memory and naïve T cells, in which 160 and 151 of 197 eGenes colocalized with GWAS loci, respectively. However, only 8, 8, and 4 eGenes showed an effect that appear to be specific to B cells, monocytes, and NK cells, respectively, while 12 eGenes showed an effect only in T cells. These observations suggest that for the vast majority of autoimmune risk loci, the effect of risk variants on gene expression level is not restricted to a single immune cell type or cell group.

We next set to understand the discrepancy between our finding that most GWAS loci impact multiple cell types and that of previous work, which suggests more cell type-specificity [36, 37]. We first analyzed colocalization status of autoimmune GWAS loci in each cell type separately, which corresponds to the general approach used by previous studies [6, 8, 14]. We found that, using this approach, the number of cell types with positive colocalization status is generally smaller—sometimes much smaller—than the number of cell types in which the eQTL effects are shared according to our analysis (Fig. 4b). We speculate that this discrepancy results from the variation in the posterior probabilities of colocalization computed by COLOC, owing to inherent noise in estimating the effect sizes and statistical significance of eQTLs (Additional File 1: Supplementary Note 4, and Additional file 2: Section 4).

We asked whether this observation was reflective of a general trend across GWAS loci. We reasoned that, under the simplifying assumption that there is only one causal eQTL at each GWAS locus, colocalized loci should show a general pattern where SNPs in high LD with the lead GWAS SNP will show strong associations with expression levels of the colocalized gene, but the eQTL associations will weaken for SNPs in lower LD. Thus, eQTLs that colocalize with a GWAS locus in all cell types should show decreasing eQTL association strength for SNPs in decreasing amount levels of LD for most or all cell types. By contrast, eQTLs that only colocalize with a GWAS locus in a single cell type, should show these patterns only in a single or a small number of cell types.

To visualize these patterns across many GWAS loci and cell types, we first found the lead GWAS SNPs at every colocalized loci and divided all SNPs within 1Mb into four bins according to their linkage disequilibrium (LD) with the lead SNP (namely, r^2 within ranges of (0, 0.25), (0.25, 0.5), (0.5, 0.75), and (0.75, 1)). Next, for each r^2 -bin, we identified the SNP with the smallest eQTL p value for the colocalized eGene in each of the 6



DICE cell groups (Fig. 4c). We then plotted the *p* values for all the colocated locus-gene pairs where the *mash* SNPs and the lead GWAS SNPs are in close LD ($r^2 > 0.8$, Fig. 4d, Figure S8a, Additional File 1: Table S5). We observed that the most significant eQTLs are often in high LD with the lead GWAS SNP for multiple cell groups (rows) when the eQTLs were determined to have shared effects. By contrast, for the eQTLs we inferred to have a cell type-specific effect, the patterns are strikingly different as the most significant eQTLs are more likely to be in lower r^2 -bins in most cell types (Fig. 4e, Figure S8b,

Additional File 1: Table S6). These findings support our high estimates of shared regulatory effects of GWAS variants across multiple cell types. These observations also suggest that COLOC is susceptible to noise in QTL mapping, especially when the sample size in QTL mapping is small. More importantly, our data indicates that previous work overestimated the fraction of GWAS loci with cell type-specific effects on gene expression levels. Indeed, we found cell type-specific colocalization in a single or two major immune cell groups for only 35 of 197 loci (17.8%), while 103 (52.3%) are eQTLs in five or more cell groups.

Limited regulatory effects specific to stimulated cells at GWAS loci

Our analysis so far indicates that about 40% of autoimmune GWAS loci have a detectable effect on gene regulation in at least one of the 18 immune cell types analyzed. We next wondered about the mechanism by which the remaining 60% of GWAS loci function. There are several possible explanations for why such a large fraction of GWAS loci do not colocalize with a regulatory QTL identified in our study. One simple explanation is that many of these GWAS loci do not impact disease risk by affecting the expression or splicing of mRNA. Instead, they may affect protein coding sequence or other as yet poorly studied molecular mechanisms, such as alternative polyadenylation [38].

To identify putative mechanisms by which trait-associated variants at uncolocalized GWAS loci function, we asked whether genes in GWAS loci without colocalization were different in terms of expression levels, enhancer density, and sequence constraint compared to those in GWAS loci with colocalization (Methods). Our analysis revealed that genes in loci without colocalization are expressed at a significantly lower level than compared to genes at loci with colocalization (Fig. 5a). In addition, we found a higher enhancer density as measured by EDS [39] (Fig. 5b), and a lower tolerance to loss-of-function mutations as measured by LOEUF [40] (Fig. 5c) for genes in uncolocalized GWAS loci.

Several studies have proposed that many autoimmune disease GWAS loci impact gene regulation in stimulated but not resting immune cells [41, 42]. Thus, it is possible that a large fraction of uncolocalized GWAS hits impact gene regulation in stimulated but not unstimulated cells. However, we found in an earlier analysis of DICE RNA-seq data that, although some exceptions exist (Figure S9), regulatory effects in stimulated CD4⁺ and CD8⁺ T cells were largely the same in unstimulated T cells. As a less direct but complementary analysis, we therefore asked whether uncolocalized GWAS loci were more likely to overlap with open chromatin regions in stimulated immune cells compared to colocalized ones using ATAC-seq data from 20 naïve and stimulated immune cells [41]. Again, we found very little support for the hypothesis that a large fraction of uncolocalized GWAS loci impact gene regulation in immune cells that were stimulated. Specifically, we observed very subtle differences in the enrichment of uncolocalized GWAS SNPs in open chromatin regions of stimulated immune cell types compared to that of colocalized GWAS SNPs (Fig. 5d, “Methods” section). When accounting for multiple testing, only 17 out of 254 tests are significant at a FDR of 5%, and the enrichment for these was modest. Thus, these analyses suggest that there are fundamental differences in the mechanisms and genes that underlie colocalized and uncolocalized autoimmune GWAS loci, but the difference cannot be simply explained by regulatory effects that are restricted to stimulated immune cells.

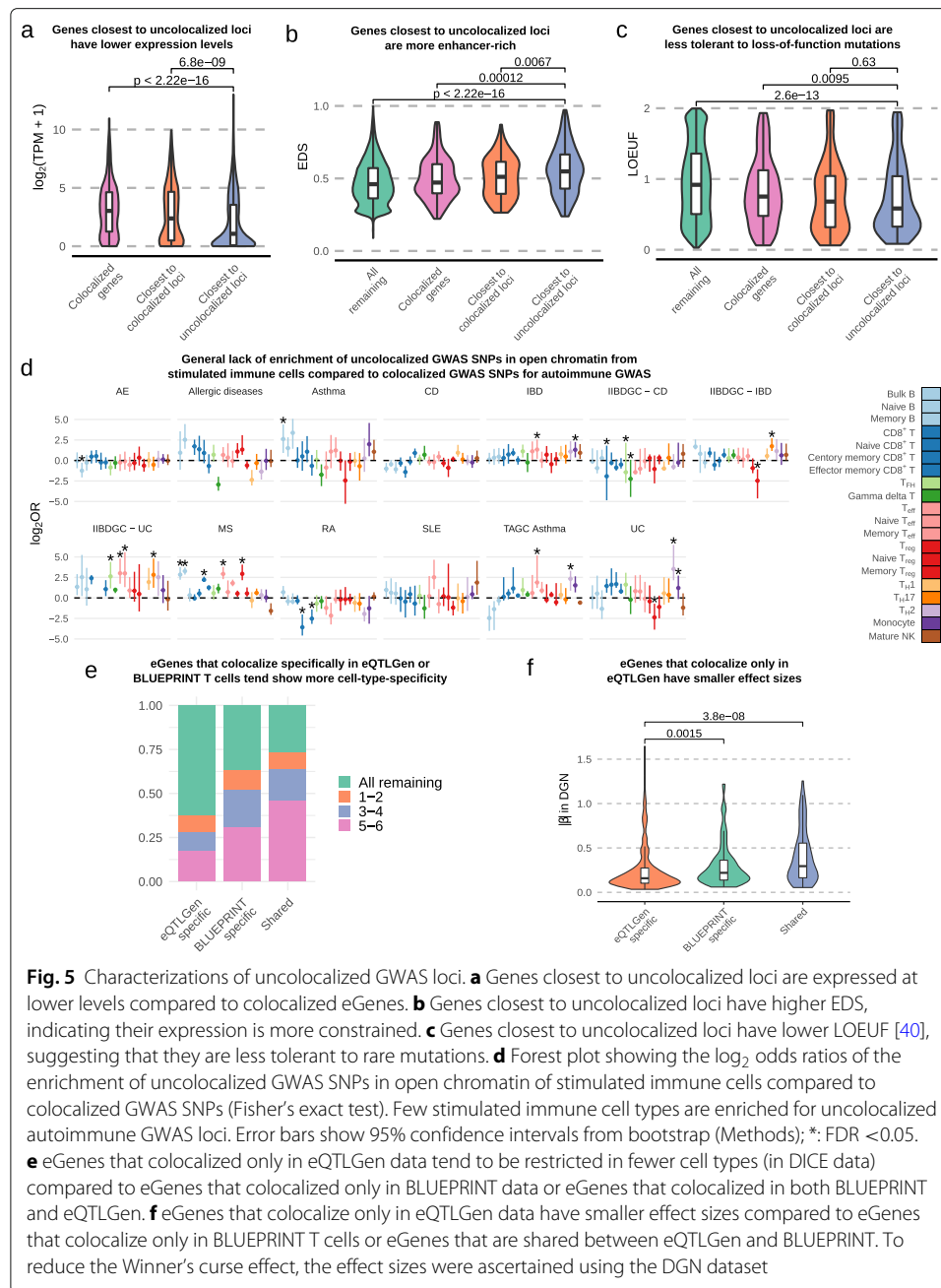


Fig. 5 Characterizations of uncolocalized GWAS loci. **a** Genes closest to uncolocalized loci are expressed at lower levels compared to colocalized eGenes. **b** Genes closest to uncolocalized loci have higher EDS, indicating their expression is more constrained. **c** Genes closest to uncolocalized loci have lower LOEUF [40], suggesting that they are less tolerant to rare mutations. **d** Forest plot showing the \log_2 odds ratios of the enrichment of uncolocalized GWAS SNPs in open chromatin of stimulated immune cells compared to colocalized GWAS SNPs (Fisher’s exact test). Few stimulated immune cell types are enriched for uncolocalized autoimmune GWAS loci. Error bars show 95% confidence intervals from bootstrap (Methods); *: FDR < 0.05. **e** eGenes that colocalized only in eQTLGen data tend to be restricted in fewer cell types (in DICE data) compared to eGenes that colocalized only in BLUEPRINT data or eGenes that colocalized in both BLUEPRINT and eQTLGen. **f** eGenes that colocalize only in eQTLGen data have smaller effect sizes compared to eGenes that colocalize only in BLUEPRINT T cells or eGenes that are shared between eQTLGen and BLUEPRINT. To reduce the Winner’s curse effect, the effect sizes were ascertained using the DGN dataset

No evidence for GWAS colocalization with small effect eQTLs at most unexplained loci

Another explanation for the large number of uncolocalized GWAS loci is that the regulatory effects of many GWAS loci are outside current range of detection owing to small sample sizes. As a simple way to test this, we performed an eQTL analysis for only the lead GWAS SNPs at uncolocalized CD GWAS loci in BLUEPRINT T cells. The smaller number of tests compared to a genome-wide analysis improved our ability to detect eQTLs with smaller effect sizes (mean absolute effect size 0.34 versus 0.64 genome-wide, Figure S10). However, we found that only a small fraction (7.97% on average) of uncolocalized autoimmune GWAS loci showed evidence of a regulatory effect using this approach. This would still leave about half of all autoimmune GWAS loci uncolocalized.

As an additional test, we asked how many uncolocalized GWAS loci could be colocalized using eQTL summary statistics from eQTLGen, which were obtained from a meta-analysis of 31,684 whole blood samples [43], including the 922 DGN samples analyzed in our study. As quality control, we first compared the eQTLGen colocalizations with that of DGN (15,269 common genes) and that of BLUEPRINT (15,373 common genes). Of the 242 autoimmune GWAS loci that colocalized with DGN eQTLs, 196 were found to replicate using the eQTLGen dataset (168 of 232 (72.4%) for BLUEPRINT). The higher replication rates for DGN was to be expected given the sample overlaps and that DGN and eQTLGen sampled the same tissue type, whole blood, while BLUEPRINT assayed sorted immune cell types.

Using eQTLGen eQTLs, we identified an additional 130 GWAS loci that colocalize in eQTLGen but not in BLUEPRINT, on average accounting for 16.8% (range: 6.6% - 35.8%) of uncolocalized loci from our BLUEPRINT analyses (Figure S11). These findings suggest that although the gain in colocalization by increasing sample size could be large for some GWAS (e.g., 35.8% for multiple sclerosis), the average increase in colocalization rate is small. As expected, colocalized eGenes specific to eQTLGen tend to not be eGenes in DICE immune cell types, or were eGenes with cell type-specificity (Fig. 5e). Additionally, the eQTLs of colocalized eGenes specific to eQTLGen have smaller effect sizes on average than that of colocalized eGenes specific to DGN, which in turn have smaller effects on average than colocalized eGenes that were identified to be shared in the DICE dataset (Fig. 5f). Thus, despite the substantial improvement in detection power afforded by the large eQTLGen sample size, the average GWAS colocalization rates for eQTLs only increased slightly, from 22.9% using BLUEPRINT compared with 29.4% using eQTLGEN. Indeed, even when colocalized loci ascertained in DGN and eQTLGen are combined together, only an average of 35.8% GWAS loci colocalized with an eQTL. While this is a relatively big increase, suggesting that the lack of colocalization at many GWAS loci is due lack of power in our eQTL analysis, these results suggest that increasing the sample size of our eQTL analysis is unlikely to account for the majority of the uncolocalized GWAS loci, at least for eQTL studies on cell types that are well-represented in whole blood.

Condition-specific profiles of H3K27ac in RA patients highlights context-dependent effects in RA pathogenesis

Finally, we hypothesized that the effects of some uncolocalized GWAS loci may be more readily interpretable in the context of the corresponding disease. While stimulation of immune cells in vitro may capture some important regulatory features reflecting disease state, we reasoned that studying immune cells sampled directly from autoimmune disease patients may better help understand the effects of uncolocalized GWAS loci. To this end, we focused specifically on rheumatoid arthritis (RA), an autoimmune disease that primarily affects synovium joints and is often associated with immune cell infiltration that leads to the build up of synovial fluid (SF) that can be collected from a joint aspiration [44].

To obtain regulatory profiles of cells in the context of RA, we first collected peripheral blood mononuclear cells (PBMC) from 6 RA patients and 4 healthy controls, as well as synovial fluid from the same RA patients. We then sorted B cells, CD4⁺ and CD8⁺ T cells, regulatory T cells and monocytes using flow cytometry (Methods), and profiled regions marked with H3K27ac using CUT&Tag (Fig. 6a). Using these data, we identified regulatory regions and quantified their activity in 5 immune cell types and 3 different

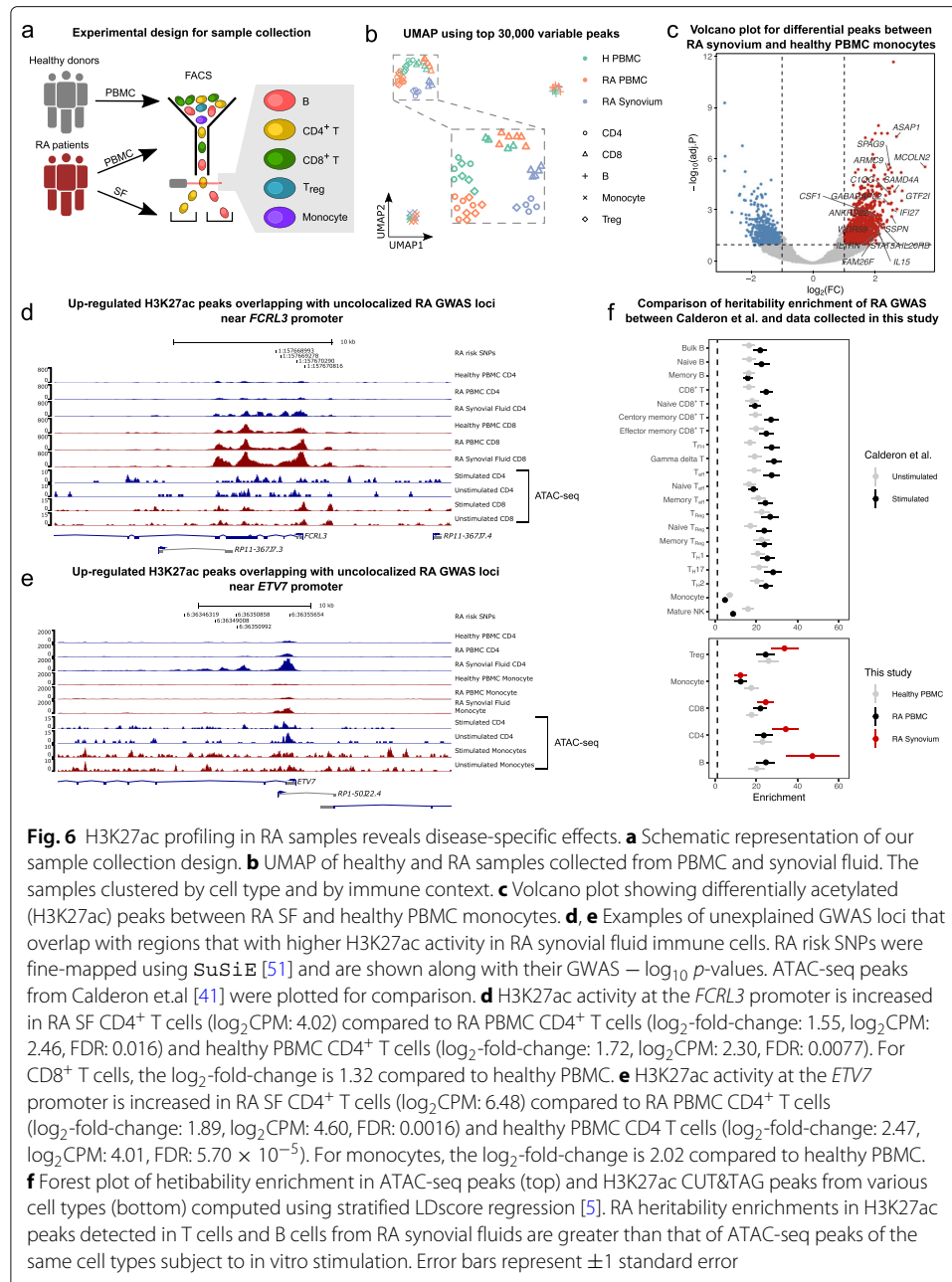


Fig. 6 H3K27ac profiling in RA samples reveals disease-specific effects. **a** Schematic representation of our sample collection design. **b** UMAP of healthy and RA samples collected from PBMC and synovial fluid. The samples clustered by cell type and by immune context. **c** Volcano plot showing differentially acetylated (H3K27ac) peaks between RA SF and healthy PBMC monocytes. **d, e** Examples of unexplained GWAS loci that overlap with regions that with higher H3K27ac activity in RA synovial fluid immune cells. RA risk SNPs were fine-mapped using SuSiE [51] and are shown along with their GWAS $-\log_{10} p$ -values. ATAC-seq peaks from Calderon et al [41] were plotted for comparison. **d** H3K27ac activity at the *FCRL3* promoter is increased in RA SF CD4⁺ T cells (\log_2 CPM: 4.02) compared to RA PBMC CD4⁺ T cells (\log_2 -fold-change: 1.55, \log_2 CPM: 2.46, FDR: 0.016) and healthy PBMC CD4⁺ T cells (\log_2 -fold-change: 1.72, \log_2 CPM: 2.30, FDR: 0.0077). For CD8⁺ T cells, the \log_2 -fold-change is 1.32 compared to healthy PBMC. **e** H3K27ac activity at the *ETV7* promoter is increased in RA SF CD4⁺ T cells (\log_2 CPM: 6.48) compared to RA PBMC CD4⁺ T cells (\log_2 -fold-change: 1.89, \log_2 CPM: 4.60, FDR: 0.0016) and healthy PBMC CD4 T cells (\log_2 -fold-change: 2.47, \log_2 CPM: 4.01, FDR: 5.70×10^{-5}). For monocytes, the \log_2 -fold-change is 2.02 compared to healthy PBMC. **f** Forest plot of heritability enrichment in ATAC-seq peaks (top) and H3K27ac CUT&TAG peaks from various cell types (bottom) computed using stratified LDscore regression [5]. RA heritability enrichments in H3K27ac peaks detected in T cells and B cells from RA synovial fluids are greater than that of ATAC-seq peaks of the same cell types subject to in vitro stimulation. Error bars represent ± 1 standard error

immune contexts corresponding to the peripheral immune context in a healthy state, the peripheral immune context in the disease state, and the immune context at the active site of inflammation. We mapped CUT&Tag 150bp paired-end reads onto the genome using Bowtie 2 [45] and identified peaks using MACS2 [46] for each sample separately. We then merged the peaks for all samples, by joining peaks that overlap, to obtain a single consensus peak set that was used for quality control and downstream analyses (Additional File 4).

As expected, UMAP visualization of the \log_2 -transformed read count-per-million (\log_2 CPM) at the top 30,000 most variable peaks in the consensus set showed separation of the major cell groups (Fig. 6b). In particular, B cells and monocytes formed

distinct clusters, while CD4⁺, CD8⁺ and regulatory T cells clustered together. Notably, cells from the same biopsy site also formed sub-clusters such that immune cells from healthy and disease PBMC clustered more closely together, while immune cells from synovial fluid clustered separately. Importantly, samples did not cluster according to batch or other technical factors (Figure S12), indicating that the observed clusters reflect biological differences between cell types and biopsy sites.

We next compared H3K27ac activity between immune cells from the different immune contexts (Methods). The general trend we observed was that H3K27ac profiles in T cells were more different between RA synovial fluid and RA PBMC than between RA PBMC and healthy PBMC. Indeed, we found 2481 and 2962 differentially acetylated peaks between RA SF and RA PBMC cells for CD4⁺ T cells and CD8⁺ T cells, respectively, compared to the 1045 and 1070 differentially acetylated peaks between RA PBMC and healthy PBMC. By contrast, the H3K27ac profile of monocytes from RA PBMC is more similar to that of RA SF monocytes than that of healthy PBMC monocytes. This finding suggests that monocytes in the peripheral blood of RA patients show similar pathogenesis signatures to synovial fluid monocytes (e.g., at the *IL1B* locus, Figure S13), and corroborates observations that were made previously using single-cell RNA-seq data [47].

We next studied the 8117 peaks that showed higher activity in immune cells from RA SF compared to immune cells from healthy PBMC. We found that many of these peaks are located near important genes that are involved in inflammation pathways and disease pathogenesis, such as *CSF1*, which modulates the differentiation of monocytes to macrophages [48], and *IL1RN* (also known as *IL1RA*), which encodes the interleukin-1 receptor antagonist protein that has been associated with autoimmune diseases including RA [49]. Interestingly, *IL1RN* expression was also found to be upregulated in monocytes treated with synovial fluid from arthritic joints [50]. Overall, we found that genes near peaks with higher activity in RA SF monocytes were enriched in functional annotations such as immune response (P value: 2.96×10^{-14} , hypergeometric-test), immune effector process (1.76×10^{-18}), and several pathways including interferon, TNF, NF- κ B, and TLR signaling pathways (1.64×10^{-03} , 5.10×10^{-05} , 3.49×10^{-03} , 8.46×10^{-03} , respectively) (Methods). Thus, the H3K27ac profiles of RA SF immune cells revealed elements that appear context-specific and relevant to RA pathogenesis.

We then asked whether differentially active peaks were enriched in unexplained GWAS loci. To answer this question, we overlapped differentially accessible peaks in all immune cells from RA patients with RA GWAS after fine-mapping using SuSiE [51] (Methods). Strikingly, we found that of the 42 uncolocalized RA GWAS loci, fine-mapped SNPs at 12 loci overlapped with a region with higher activity in RA immune cells (6 loci for healthy PBMC, bootstrap p -value 0.026, Methods). For example, we found that a lead GWAS SNP lies within a differentially active peak at the promoter region of *FCRL3* in CD4⁺ and CD8⁺ T cells (FDR: 7.7×10^{-3} for CD4⁺ and 2.6×10^{-2} for CD8⁺ T cells, Fig. 6d). In another example, the RA lead SNP overlaps with an H3K27ac peak located near the promoter of *ETV7* which showed higher activity in both RA SF CD4⁺ T cells compared to the respective cell types from RA PBMC (FDR: 1.6×10^{-3}) and healthy PBMC (FDR: 5.7×10^{-5}). The activity of this regulatory region was also higher in RA SF monocytes compared to healthy PBMC monocytes (FDR: 4.5×10^{-3} , Fig. 6e).

To further assess the relevance of each immune context on the study of disease etiology, we quantified the enrichment of RA heritability in H3K27ac peaks identified in the different immune contexts using stratified LDscore regression [5]. To establish a baseline for comparison, we used accessible chromatin regions identified using ATAC-seq data from unstimulated and stimulated immune cell types [41] (Methods). Our estimates recapitulated the findings from the original study [41] in which CD8⁺ T cells and delta gamma T cells showed the largest increase in heritability enrichment subsequent to stimulation (~30-fold vs ~20-fold enrichments for stimulated versus unstimulated). We then applied the same analysis using our H3K27ac peaks. We found that while the estimated RA heritability was similarly enriched in ATAC-seq peaks from unstimulated immune cells and in H3K27ac peaks from RA PBMC and Healthy PMBC (~20-fold), the heritability enrichment was greater in H3K27ac peaks from RA SF B cells, CD4⁺ T cells, and Tregs than in ATAC-seq peaks from in vitro stimulation of the same cell types (Fig. 6f).

Our analyses therefore show that there are significant differences in the regulatory landscapes of immune cell types across disease states and immune contexts. In particular, we found that the regulatory landscape of cells extracted from the active site of RA inflammation showed striking differences when compared to that of circulating immune cells in the periphery of both RA patients and healthy individuals. Importantly, we find that the regulatory regions identified in immune cells from RA synovial fluid overlap with many uncolocalized GWAS loci and are the most highly enriched in RA SNP heritability. Altogether, these observations indicate the importance of studying cell types in the correct disease context in order to elucidate the genetic etiology of a disease.

Discussion

The goal of this study was to establish a detailed accounting of the effects of genetic variants on gene regulation in immune cells and their overlap with genetic effects on human traits and disease. Recent studies suggested that fewer than a third of GWAS loci colocalize with an eQTL [14, 15]. This finding implies that much is left to be understood about the mechanisms by which genetic variants impact human traits.

There are several possible explanations for the small fraction of GWAS loci that colocalize with an expression QTLs. Our work evaluated the possibilities that (i) there exist genetic effects on gene regulation other than steady state gene expression levels, (ii) genetic effects are often restricted to cell types and cell states that are causal for the trait, and (iii) genetic effects are often too small to be detected, even in the causal cell types or cell states. These possibilities are not mutually exclusive, but the implications are different for how we should design future human genomics research. For example, if trait-associated variants often impact mRNA splicing but not steady state mRNA expression, then a more widespread focus on mapping the effects of genetic variants on mRNA splicing is needed. If most disease-associated genetic effects are very specific to cell types and cell states that are relevant to the trait, then studying eQTLs identified in bulk, unsorted, tissues will have limited success in elucidating the mechanisms underlying most GWAS loci.

Using our harmonized regulatory QTL data, we found that eQTLs and sQTLs together colocalized with up to 45% of trait-associated loci for the 72 GWAS we analyzed. On average, 40.4% of significant loci from the 50 immune-related GWAS colocalized with a regulatory QTLs, a larger proportion compared to an average of 26.4% for the 21

non-immune GWAS we analyzed (excluding Alzheimer's disease, 55.2% colocalized). One of the caveats in our colocalization analysis is the use of the method COLOC. Although COLOC is a very popular method for colocalization analyses, it uses priors that, when altered, can impact substantially the computed posterior probabilities that the causal eQTL and GWAS variants are the same variants. Reassuringly, when we used another colocalization method, HyPrColoc [52], that does not rely on user-defined priors, we were able to replicate nearly all colocalized genes identified using COLOC, indicating that our colocalization analyses are robust and replicable (Additional File 1: Supplementary Note 4).

Our data also allowed us to ask whether regulatory QTLs are likely to be active in many immune cell types, or only in few or a single cell type. We found that at least one third of eQTLs (81% of sQTLs) are shared across all 15 immune cell types we analyzed from the DICE dataset. For closely-related cell types, we found that the fraction of shared eQTLs was as high as 96% (99% for sQTLs). Intriguingly, activated and naïve T cells share nearly 70% of detected eGenes. Thus, QTL effects appear similar across many cell types and cell states. One important implication of this finding is that eQTLs that colocalize with GWAS SNPs in one cell type are also likely to be active in other cell types. Thus, eQTLs that colocalize at a GWAS locus in one cell type should, in general, colocalize in the other cell types. Indeed, after accounting for variability in the posterior probabilities of colocalization reported by COLOC owing to the inherent noise in QTL mapping, we found that the majority of GWAS loci colocalizes with the same QTL in multiple cell types. Altogether, these data questions the notion that the vast majority GWAS SNPs affect gene regulation in a very cell type-specific manner as highlighted in several studies [8, 12]. Thus, the use of regulatory QTLs from proxy cell types or tissues, e.g., from the GTEx consortium, to identify causal genes may be well justified for a large fraction of GWAS loci.

A noteworthy finding from our colocalization analysis is that genetic variants that impact mRNA splicing often colocalize with a GWAS signal. Indeed, if we considered eQTLs only, our rates of colocalization would be very similar to that of previous studies (26.2% vs 21%) [14]. Instead, when sQTLs were tested for colocalization, we found that more GWAS loci colocalized with sQTLs than with eQTLs. It is worth noting however, that a substantial number of GWAS loci colocalized with both an eQTL and an sQTL. This may be due to horizontal pleiotropy, whereby a genetic locus can influence the expression level of a gene, as well as the splicing of an intron in the same or a different gene. Another possible explanation for this observation is that eQTL effects are often mediated by sQTLs or vice-versa. A colocalization analysis for sQTLs conditioned on the eQTLs would be necessary to tease apart these possibilities but is outside the scope of our work.

Despite a substantial increase in colocalization rates in our study, we find that for most traits, over half of all GWAS loci do not colocalize with a regulatory QTL. Interestingly, we found several differences between genes at colocalized GWAS loci and those at uncolocalized loci. Genes at GWAS loci without colocalized regulatory QTLs tend to be more lowly expressed, have higher enhancer density, and are less tolerant to loss-of-function mutations. These findings suggest that genes at uncolocalized GWAS loci may be subject to stronger constraints both at the levels of gene regulation and sequence conservation. Thus, a plausible explanation is that genetic effects at these loci are on average smaller and more cell type or context-specific compared to genes at GWAS loci with colocalization.

This hypothesis is consistent with the idea that much larger sample sizes may be required to find the causal QTL effects that explain the associations at GWAS loci without colocalization. That said, our colocalization analyses on QTL datasets with very large sample sizes (DGN: $N = 900$, eQTLGen: $N = 31,684$) revealed that the rates of colocalization only increased slightly despite the large increase in our power to detect low-effect QTLs. We speculate that an important reason for the modest increase in colocalization is because both DGN and eQTLGen QTL data are from whole blood samples, which are less likely to capture genetic effects that are cell type or context-specific.

One intriguing finding from our analyses is that eQTLs identified specifically in *in vitro* stimulated immune cells from DICE colocalized with only a small number of GWAS loci that did not colocalize with QTLs from unstimulated cells. This observation might seem surprising because a recent paper showed that autoimmune disease SNP heritability is more highly enriched in accessible chromatin from *in vitro* stimulated immune cells compared to naïve immune cells [41]. However, we should note that our data does indeed suggest that SNPs at colocalized and uncolocalized GWAS loci are more highly enriched in open chromatin from stimulated cells compared to unstimulated cells. The differences in the enrichment, however, is negligible, suggesting that stimulation-specific effects can not explain why a large fraction of GWAS loci do not colocalize with the regulatory QTLs identified in our study.

One possible explanation for the modest increase in the colocalization rates, when using eQTLs identified in stimulated immune cells, is that the immune cells stimulated *in vitro* only partly recapitulate gene regulation in the *in vivo* disease context. Thus, although many regulatory elements are primed to be activated subsequent to *in vitro* stimuli – thereby capturing some of the important regulatory regions relevant to disease – they may require additional factors to fully capture the effects of genetic variants on gene expression levels in the disease context. In support of this, [53] found that stimulating immune cells *in vitro* was able to recapitulate gene expression signatures of immune cells from rheumatoid arthritis (RA) patients when 6 different cytokines were used together, but not when the cytokines were used on their own.

To better understand the role of context on our ability to interpret GWAS signals, we collected H3K27ac measurements in healthy and RA patients using CUT&Tag to use as proxy for enhancer and promoter activity. Although the sample sizes are too small for a QTL analysis, we were able to use these data to ask whether gene regulatory data in the disease context could aid us to identify putative mechanisms that underlie RA GWAS hits, in particular for loci with no QTL colocalization. We found that SNPs at 12 out of 42 uncolocalized GWAS loci overlap with regions with increased H3K27ac levels in immune cells from RA synovial fluid. Remarkably, we also found that regions marked by H3K27ac in immune cells from RA synovial fluid were more highly enriched in RA heritability than compared to healthy or RA immune cells collected from peripheral blood. Additionally, our initial analyses suggest that the RA GWAS heritability enrichments in regulatory regions identified in RA synovial fluid immune cells are even higher than in that of *in vitro* stimulated immune cells. We should note here that caution must be used when interpreting these results as the data type collected in these two studies differ (ATAC-seq versus CUT&Tag). Nevertheless, these preliminary analyses indicate that studying the regulatory effects of genetic variants in the disease context may be critical for discovering the mechanisms behind a large number of GWAS loci without colocalization.

Methods

Data processing

To harmonize the set of genetic variants across all four datasets, we imputed the genotypes of all individuals in the four studies using the 1000G Phase 3 v5 as a common reference panel (Michigan Imputation Server [54]). Following imputation, only non-duplicated genetic variants with INFO score larger than 0.9 were retained. We filtered variants with Hardy-Weinberg Equilibrium (HWE) p values below 10^{-5} , with missing genotype rate higher than 5%, and with minor allele frequency below 5% using PLINK v1.9 [55]. We used the remaining set of variants in all subsequent analyses unless otherwise noted. To exclude outlier individuals, we calculated genotype principal components (PCs) using `smartpca` [56]. Five outliers in the DICE dataset were identified and removed from downstream analyses.

To quantify gene expression levels, we used `Kallisto` [57] and summed the transcript per million (TPM) estimates of all GENCODE 19 [58] isoforms to obtain a gene-level TPM. The gene-level TPM were then scaled and quantile-quantile normalized as described before [17]. Gene expression principal components were calculated using the `prcomp` function in R. To quantify RNA splicing, RNA-seq reads were aligned to the hg19 reference genome using `STAR 2.6.0` [59] with the GENCODE 19 annotation. To avoid reads mapping with allelic bias, we used `WASP` [60] as implemented in `STAR 2.6.0` by providing the corresponding genotype data. This is an important step as we found a substantial increase in the number of false positive splicing QTL due to allelic bias in read mapping. Indeed, when reads representing different alleles map to different regions of the genome, QTL mapping will be susceptible to identifying spurious associations between the alleles and read coverage at those genomic regions [23]. Exon-exon junctions were extracted using `RegTools` [61], and clustered and quantified using `LeafCutter` [23]. As expected, we observed that the number of exon-exon junctions identified in each sample is positively correlated with the sequencing depth in the DICE consortium (Figure S1). To harmonize quantification for splicing junction usage across cell types and datasets in all 18 immune cell types, clusters were merged and the merged union was used to re-calculate intron usage in all samples.

MashR analysis in the DICE dataset

To quantify the sharing of eQTLs and sQTLs in the DICE dataset, we followed the workflow provided by the authors of `MashR` (<https://github.com/stephenslab/gtexresults>) that was previously described in [19]. Briefly, standard errors of QTL effect sizes were calculated from `FastQTL` nominal output, which were used together with effect sizes as the input for `mash`. To quantify the correlation structure of the null tests, 30% of all tests were randomly sampled (referred to as the “random” set). To obtain a confident set of QTLs for each feature (gene or intron), the SNP with the smallest P -value across all tested SNPs and all cell types were extracted for each feature. This resulted in a feature-by-sample matrix of effect sizes and their standard errors without missing values referred to as the “strong” set. For eQTLs, we included all protein coding genes. For sQTLs, we included all introns. Data-driven covariance matrices were calculated from the “strong” set. We then built a `mash` model using the “random” set with the exchange effects (EE) mode to estimate the priors. This model was then applied to the “strong” set to calculate the posterior mean effect sizes (`mash` effect sizes). Significant QTLs after `mash` analysis were

feature-SNP pairs with local false sign rate (LFSR) below 0.05, as suggested by [19]. The level of QTLs sharing was quantified as both overall sharing and pairwise sharing. Overall, sharing was determined to be the number of cell types in which a given feature has a regulatory QTL (LFSR < 0.05). Pairwise sharing was quantified both by magnitude and by sign. Share-by-magnitude between two cell types correspond to the proportion of QTLs that is significant in one of the cell types and posterior mean effect sizes differ by no more than twofold. Share-by-sign between two cell types correspond to the proportion of QTLs that was significant in one of the cell types and had the same sign. The 15 cell types in DICE were grouped into 6 cell groups based on the eQTL sharing-by-magnitude (see Fig. 2b).

Characterization of regulatory QTLs

To calculate the distance between eQTLs and their target genes, we defined the promoter of each gene as the region 2000 bp upstream and 500 bp downstream of TSS. We tested the enrichment of eQTLs in regulatory elements from Ensembl Regulatory Build and consensus ATAC-seq peak set from Calderon et al. [41]. We categorized all ATAC-seq peaks to be either an enhancer or a promoter based on whether they overlap with any promoter region (2000 bp upstream and 500 bp downstream of TSS). The observed and expected number of QTLs overlapping with each feature was estimated using the `fenrich` command from `QTLtools` [62], and the odds ratios of enrichment were calculated by supplying those number to Fisher's exact test in R. We validated eQTLs from DICE in other datasets using π_1 statistics [63], stratifying eQTLs by their levels of sharing across six cell groups estimated by *mash* (specific: in one cell group; intermediate: 2–5 cell groups; shared: 6 cell groups). The 95% confidence intervals of π_1 was estimated using 1000 bootstraps (i.e., re-sampling DICE eQTLs with replacement).

Colocalization

COLOC Colocalization analyses were performed between eQTLs/sQTLs and 72 publicly available GWAS summary statistics for 11 autoimmune diseases (14 studies), namely, rheumatoid arthritis (RA) [64], Crohn's disease (CD) [27, 30], ulcerative colitis (UC) [27, 30], inflammatory bowel disease (IBD) [27, 30], allergy and eczema (AE) [65], asthma, hay fever and eczema (allergy for short) [66], apoptotic dermatitis (ApD) [67], asthma [68, 69], systemic lupus erythematosus (SLE) [70] and multiple sclerosis [71]. We also collected 36 GWAS for blood-related traits [72], 11 GWAS related to heart functions and circulation system [73], and several other traits including type 2 diabetes (T2D) [74], Alzheimer's disease (AD) [75], Parkinson's disease (PD) [76], estimated glomerular filtration rate (eGFR) [77], height [78], and breast cancer survival [79] and other cancers/neoplasms [73]. We considered the 14 autoimmune and the 36 blood-related GWAS as immune GWAS, and the rest 22 GWAS as non-immune GWAS.

To assess colocalization between GWAS loci and QTLs, we first identified the lead GWAS variants and their flanking region in which colocalization was to be tested. Specifically, all variants available in the GWAS summary statistics were sorted by p -values in increasing order. Starting from the variant with the smallest p -value (lead variant), variants within the 500 Kb window on either side of the leading variant were removed. This resulted in a 1Mbp GWAS locus for colocalization analysis. The same procedure was then applied to the next most significant variant among the remaining variants, until no

variant with p value below 10^{-7} was left. The HLA region (Chr6: 25–35 Mb) was excluded from colocalization. Only GWAS with more than 10 identified loci were included in our analysis. For each GWAS locus identified above, colocalization was tested only if it harbored a regulatory QTL with beta-distribution permuted p value below 0.01 ($\text{bpval} < 0.01$) as reported by *FastQTL* in the 1 Mb window flanking that leading GWAS SNP. Default priors were used for *COLOC*. We set $\text{PP4} > 0.75$ as the threshold for colocalization. The colocalization proportion was calculated as the proportion of colocalized loci among all identified loci in a GWAS.

Colocalization results were visualized using a function adapted from *LocusCompare* [80]. For a given locus, SNP with the largest posterior probability from *COLOC* was defined as the colocalized SNP. r^2 relative to the colocalized SNP were calculated from the genotypes in the QTL study. To visualize the sQTL in the form of a Sashimi plot [81], we first grouped individuals by their genotypes, and then extracted RNA-seq reads that mapped to the cluster that contains the intron to be visualized. To make the coverage comparable between different genotypes, we scaled the read coverage by the number of individuals that carry each genotype using the `scaleFactor` argument in `bamCoverage` from *DeepTools* [82] when generating bigWig files. The coverage was then visualized using *pyGenomeTracks* [83].

Cis-eQTL data of *eQTLGen* [43] was directly obtained from the website (<https://eqtlgen.org/cis-eqtls.html>). We also downloaded allele frequencies from 26,609 *eQTLGen* samples (excluding Framingham Heart Study), which were used in our colocalization analysis. Of note, the DGN dataset is also included in *eQTLGen* meta-analysis, but does not alter the interpretation of any of our analyses.

HyPrColoc The GWAS-gene pairs tested in *HyPrColoc* were selected in the same way as *COLOC*. We set $\text{PP} > 0.25$ as the threshold for colocalization as recommended by the authors [52].

Validation of immune-cell-specific colocalization for non-immune traits We validated colocalization of 14 non-immune traits (11 heart-related, AD, PD and breast cancer survival) in DICE immune cells using the GTEx V7 eQTLs. We first chose several tissues in GTEx that are most relevant to each GWAS trait. For heart-related traits, we chose tissues in heart and circulation system (Artery - Aorta, Artery - Coronary, Artery - Tibial, Heart - Atrial Appendage, Heart - Left Ventricle). For AD and PD, we included the 13 brain tissues (Brain - Amygdala, Brain - Anterior cingulate cortex (BA24), Brain - Caudate (basal ganglia), Brain - Cerebellar Hemisphere, Brain - Cerebellum, Brain - Cortex, Brain - Frontal Cortex (BA9), Brain - Hippocampus, Brain - Hypothalamus, Brain - Nucleus accumbens (basal ganglia), Brain - Putamen (basal ganglia), Brain - Spinal cord (cervical c-1), Brain - Substantia nigra). For breast cancer survival, we used adipose tissues and breast tissue (Adipose - Subcutaneous, Adipose - Visceral (Omentum), Breast - Mammary Tissue). We then identified all the colocalized gene-SNP pairs for these 14 GWAS in DICE, and extracted their P values from GTEx eQTLs in the relevant tissues, as well as from DICE eQTLs in all immune cell types. Given that a large proportion of eQTLs are shared in DICE, we grouped the 15 immune cell types into 6 groups, assigning the smallest P value from all cell types within a given group to that group for each gene. We used Bonferroni correction to adjust P values for multiple testing. Finally, we calculated

the proportion gene-SNP pair that has adjusted P value below 0.05 in DICE but not GTEx tissues.

Characterizations of uncolocalized GWAS loci We restricted this analysis to the loci from the 14 autoimmune GWAS that did not colocalize with a in BLUEPRINT QTL. All genes were classified into four categories: genes with an eQTL that colocalized at a GWAS locus, genes that are the closest to a GWAS locus, genes that are closest to an uncolocalized GWAS locus, and all remaining genes. We compared gene expression level in the three BLUEPRINT cell types separately. The gene expression level values for the three cell types were combined and plotted in Fig. 5a. We also obtained Enhancer-domain score (EDS) [39] and “loss-of-function observed/expected upper bound fraction” (LOEUF) [40] for all available genes and compared the distribution of EDS and LOEUF across the four categories above.

To test the enrichment of uncolocalized loci in ATAC-seq peaks in stimulated immune cells, we constructed a contingency table by counting the number of colocalized and uncolocalized loci overlapping stimulated and unstimulated ATAC-seq peaks, respectively. We then tested the hypothesis that uncolocalized loci were more highly enriched in stimulated open chromatin regions compared to colocalized loci using Fisher’s exact test. We estimated 95% confidential interval of estimates by bootstrapping uncolocalized GWAS loci 1000 times with replacement.

We reasoned that regulatory effects of many uncolocalized GWAS loci might be too small to be detected due to small sample sizes. To test this possibility, we ascertained eQTLs only at uncolocalized GWAS loci. Briefly, we extracted QTL tests at lead SNP of uncolocalized loci. GWAS locus-gene pairs that have already been tested in COLOC but did not colocalize were filtered. Since it is common for one lead SNP to be associated with many genes, we adjusted the P values by number of tested genes at each loci using Bonferroni correction and picked the gene with the smallest P value. We then calculated the proportion of genes with P value below 0.05. This analysis was applied to each autoimmune GWAS in each cell type in BLUEPRINT dataset.

RA samples collection and analysis

Sample collection and CUT&Tag experiment All of the clinical samples were obtained from Xijing Hospital. Peripheral blood and synovial fluid samples were collected from 6 RA patients at the Department of Clinical Immunology, Xijing Hospital. All of the RA patients fulfilled the 1987 revised American College of Rheumatology criteria and the ACR 2010 Rheumatoid Arthritis classification criteria [84], and their clinical characteristics are shown in Additional File 1: Table S7. In addition, peripheral blood samples were gathered from 4 healthy individuals. All blood and synovial fluid samples were subjected to gradient centrifugation using lymphocyte separation medium (MP Biomedicals, 0850494) to isolate mononuclear cells, which were cryopreserved for later experiments.

The cryopreserved mononuclear cells were thawed into RPMI/10%FBS, washed once in sterile phosphate-buffered saline (PBS; Beyotime, ST476), and stained with the following antibodies in PBS for 30 min: anti-CD3-APC/Cy7 (Biolegend, 300426), anti-CD4-PE/Cy7 (Biolegend, 357410), anti-CD8-Percp/Cy5.5 (Biolegend, 301032), anti-CD25-PE/CF594(BD Horizon,562525), anti-CD19-FITC (Biolegend,302206), and anti-CD14-APC (Biolegend, 301808). CD4⁺ T cells (CD3⁺, CD4⁺, CD8⁻), CD8⁺ T cells(CD3⁺,

CD4⁻, CD8⁺), T_{reg} cells (CD3⁺, CD4⁺, CD8⁻, CD25⁺), B cells (CD3⁻, CD19⁺), and monocytes (CD3⁻, CD14⁺) were sorted by FACSaria III (BD Pharmingen, San Diego, USA) directly into wash buffer for CUT&Tag, with a maximum of 1×10^5 cells for each cell type. We profiled H3K27ac (abcam ab4729) for each cell type following standard CUT&Tag protocol (<https://www.protocols.io/view/bench-top-cut-amp-tag-z6hf9b6>) [21]. Samples were processed in different batches, and we ensured to include at least one healthy individual and one RA patient in each batch to minimize batch effects that align with biological differences that we are interested in.

CUT&Tag data analysis The DNA libraries were subjected to 150bp paired-end (PE) sequencing. Sequencing reads were aligned to human reference genome hg19 using Bowtie 2 [45] with parameters `-local -very-sensitive-local -no-unal -no-mixed -no-discordant -phred33 -minins 10 -maxins 700`. Aligned reads were filtered using Samtools with `-F 1804 -f 2 -q 30` [85]. Samples with fewer than 2M reads were excluded from subsequent analyses. Filtered BAM files for samples that have the same disease status (healthy/RA), tissue-type (PBMC/SF) and cell type were merged. Read coverage was calculated using `bamCoverage` in 10bp window normalized by RPKM [82]. H3K27ac peaks were called from the merged BAM files using MACS2 with parameters `-format BAMPE -broad -broad-cutoff 0.1 -qvalue 0.1 -extsize 146` [46]. We reasoned that calling peaks from merged BAM files increases the signal-to-noise ratio. To generate a consensus peak set, we merged all the peaks using `bedtools merge` [86], resulting in 90,412 peaks. We then counted the number of fragments overlapping with the consensus peak set in each sample using `featureCounts` [87].

Differential peak analysis was performed using `limma` [88]. We calculated average \log_2 CPM across samples with the same disease status, tissue-type, and cell type. This average \log_2 CPM was only used to filter our peaks with low fragments counts. Peaks with average \log_2 CPM below 2 in all groups were excluded from differential analysis. Then, normalization factors were calculated from the remaining peaks using the TMM method, and counts in each sample converted to \log_2 CPM. Since samples were processed in different batches, we used `ComBat` to adjust for batches while including disease status, tissue-type, and cell type as our variable-of-interest. We constructed a contrast matrix comparing RA SF vs. RA PBMC, RA SF vs. Healthy PBMC, and RA PBMC vs. Healthy PBMC in each, and applied the `trend` method. Differential peaks were defined as \log_2 -fold-change (\log_2 (FC)) larger than 1 or smaller than -1, and FDR below 0.1.

We overlapped H3K27ac peaks up-regulated in RA samples with uncolocalized RA GWAS loci. We first fine-mapped RA GWAS summary statistics using `Susie` [51]. Fine-mapping was performed at each locus we used in our colocalization analysis. We supplied GWAS Z-scores, genotype correlation matrix from CEU and GBR from the 1000 Genome Project as the reference panel and the sample size of reference panel to the `susie_rss` function.

We estimated the enrichment of RA SNP heritability in our H3K27ac peaks using Stratified LD Score Regression (S-LDSC) [5]. We used MACS2 peaks from merged BAM files, which were extended by 500 bp on both sides. To reproduce the heritability analysis from Calderon et al. [41], we used the MACS2 peaks shared by the authors.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-021-02334-x>.

Additional file 1: A master file for Table S1-8.

Additional file 2: Supplementary Notes and Supplementary Figures.

Additional file 3: A table for all colocalization results.

Additional file 4: The consensus peak set with fragment counts from H3K27ac CUT&Tag experiments.

Additional file 5: Review history

Acknowledgements

We thank D Calderon, BE Mittleman, A Shah, F Morgante, and W Lin for comments on the manuscript.

Peer review information

Anahita Bishop was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Review history

The review history is available as Additional file 5.

Authors' contributions

Y.L.L. conceived of the project. Z.M. performed computational analyses. W.W. performed the experiments with the help of B.F. Y.L.L. and Z.M. wrote the manuscript with contribution from W.W. P.Z. and J.M. collected the human samples. All author(s) read and approved the final manuscript.

Funding

This work was supported by the US National Institutes of Health (R01GM130738 to YI Li, R01CA242929 to Z Mu). This work was completed in part with resources provided by the University of Chicago Research Computing Center. This work was also supported by Shaanxi Province Natural Science Foundation Research Program (directed entrustment Project) 2019ZY-CXPT-03-01 to P. Zhu.

Availability of data and materials

All the processed data are available through Zenodo [22]. Raw sequencing files for H3K27ac CUT&Tag experiments are available through EGA Archive upon request (EGAD00001007003) [89]. A brief description for each data file can be found in Additional File 1: Table S8. Data from Calderon et al. [41] can be found at http://web.stanford.edu/group/pritchardlab/dataArchive/immune_atlas_web/index.html. eQTLGen [43] summary statistics can be found at <https://eqtlgen.org/cis-eqtls.html>, gnomAD LOEUF [40] and Enhancer-domain score (EDS) [39] can be found at <https://gnomad.broadinstitute.org/downloads> and <https://doi.org/10.1016/j.ajhg.2020.01.012>, respectively.

Declarations

Ethics approval and consent to participate

The collection of human samples was approved by the Institutional Review Board of the Xijing Hospital, Xi'an, China (KY-20192006-F-1) and has obtained written informed consent from all participants. All the experimental methods complied with the Helsinki Declaration.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Committee on Genetics, Genomics & Systems Biology, University of Chicago, Chicago, IL, USA. ²Department of Clinical Immunology, Xijing Hospital, Xi'an, China. ³National Translational Science Center for Molecular Medicine, Xi'an, China. ⁴Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL, USA. ⁵Department of Human Genetics, Department of Medicine, University of Chicago, Chicago, IL, USA.

Received: 5 September 2020 Accepted: 30 March 2021

Published online: 29 April 2021

References

1. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 10 years of GWAS discovery: biology, function, and translation. *Am J Hum Genet.* 2017;101(1):5–22.
2. Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci.* 2009;106(23):9362–7.
3. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* 2010;6(4):1000888.
4. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science.* 2012;337(6099):1190–5.

5. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, Anttila V, Xu H, Zang C, Farh K, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet.* 2015;47(11):1228.
6. The GTEx Consortium. The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science.* 2020;369(6509):1318–30.
7. Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martín D, Watt S, Yan Y, Kundu K, Ecker S, et al. Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell.* 2016;167(5):1398–414.
8. Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Zapardiel-Gonzalo J, Ha B, Altay G, Greenbaum JA, McVicker G, et al. Impact of genetic polymorphisms on human immune cell gene expression. *Cell.* 2018;175(6):1701–15.
9. Lappalainen T, Sammeth M, Friedländer MR, AC't Hoen P, Monlong J, Rivas MA, Gonzalez-Porta M, Kurbatova N, Griebel T, Ferreira PG, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature.* 2013;501(7468):506.
10. Battle A, Mostafavi S, Zhu X, Potash JB, Weissman MM, McCormick C, Haudenschild CD, Beckman KB, Shi J, Mei R, et al. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res.* 2014;24(1):14–24.
11. Nédélec Y, Sanz J, Baharian G, Szpiech ZA, Pacis A, Dumaine A, Grenier J-C, Freiman A, Sams AJ, Hebert S, et al. Genetic ancestry and natural selection drive population differences in immune responses to pathogens. *Cell.* 2016;167(3):657–69.
12. Alasoo K, Rodrigues J, Mukhopadhyay S, Knights AJ, Mann AL, Kundu K, Hale C, Dougan G, Gaffney DJ. Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nat Genet.* 2018;50(3):424–31.
13. Kim-Hellmuth S, Bechheim M, Pütz B, Mohammadi P, Nédélec Y, Giangreco N, Becker J, Kaiser V, Fricker N, Beier E, et al. Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nat Commun.* 2017;8(1):266.
14. Chun S, Casparino A, Patsopoulos NA, Croteau-Chonka DC, Raby BA, De Jager PL, Sunyaev SR, Cotsapas C. Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nat Genet.* 2017;49(4):600.
15. Consortium G, et al. Genetic effects on gene expression across human tissues. *Nature.* 2017;550(7675):204–13.
16. Yao DW, O'connor LJ, Price AL, Gusev A. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nat Genet.* 2020:1–8.
17. Li Yi, Van De Geijn B, Raj A, Knowles DA, Petti AA, Golan D, Gilad Y, Pritchard JK. RNA splicing is a primary link between genetic variation and disease. *Science.* 2016;352(6285):600–4.
18. Franzén O, Ermel R, Cohain A, Akers NK, Di Narzo A, Talukdar HA, Foroughi-Asl H, Giambartolomei C, Fullard JF, Sukhvasi K, et al. Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science.* 2016;353(6301):827–30.
19. Urbut SM, Wang G, Carbonetto P, Stephens M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat Genet.* 2019;51(1):187–95.
20. Qi T, Wu Y, Zeng J, Zhang F, Xue A, Jiang L, Zhu Z, Kemper K, Yengo L, Zheng Z, et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat Commun.* 2018;9(1):1–12.
21. Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, Ahmad K, Henikoff S. Cut&tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun.* 2019;10(1):1–10.
22. Mu Z, Wei W, Fair BJ, Miao J, Zhu P, Li Y. Impact of cell type and context-dependent regulatory variants on human immune traits. *Datasets.* Zenodo. 2021. <https://doi.org/10.5281/zenodo.4480206>.
23. Li Yi, Knowles DA, Humphrey J, Barbeira AN, Dickinson SP, Im HK, Pritchard JK. Annotation-free quantification of rna splicing using leafcutter. *Nat Genet.* 2018;50(1):151.
24. Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics.* 2015;32(10):1479–85.
25. Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Arcelus MG, Sekowska M, et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science.* 2009;325(5945):1246–50.
26. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. Bayesian test for localisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 2014;10(5):1004383.
27. de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, Jostins L, Rice DL, Gutierrez-Achury J, Ji S-G, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet.* 2017;49(2):256.
28. Maeda S, Hsu L-C, Liu H, Bankston LA, Iimura M, Kagnoff MF, Eckmann L, Karin M. Nod2 mutation in crohn's disease potentiates NF- κ B activity and il-1 β processing. *Science.* 2005;307(5710):734–8.
29. Ghosh S, Goldin E, Gordon FH, Malchow HA, Rask-Madsen J, Rutgeerts P, Vyhňálek P, Zádorová Z, Palmer T, Donoghue S. Natalizumab for active crohn's disease. *N Engl J Med.* 2003;348(1):24–32.
30. Liu JZ, Van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, Ripke S, Lee JC, Jostins L, Shah T, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet.* 2015;47(9):979.
31. Bloch DB, Suzanne M, Guigaouri P, Filippov A, Bloch KD. Identification and characterization of a leukocyte-specific component of the nuclear body. *J Biol Chem.* 1996;271(46):29198–204.
32. Mehta S, Cronkite DA, Basavappa M, Saunders TL, Adiliaghdam F, Amatullah H, Morrison SA, Pagan JD, Anthony RM, Tonnerre P, et al. Maintenance of macrophage transcriptional programs and intestinal homeostasis by epigenetic reader SP140. *Sci Immunol.* 2017;2(9):

33. Karaky M, Fedetz M, Potenciano V, Andrés-León E, Codina AE, Barrionuevo C, Alcina A, Matesanz F. Sp140 regulates the expression of immune-related genes associated with multiple sclerosis and other autoimmune diseases by NF- κ B inhibition. *Hum Mol Genet.* 2018;27(23):4012–23.
34. Matesanz F, Potenciano V, Fedetz M, Ramos-Mozo P, Abad-Grau MdM, Karaky M, Barrionuevo C, Izquierdo G, Ruiz-Peña JL, García-Sánchez MI, et al. A functional variant that affects exon-skipping and protein expression of SP140 as genetic mechanism predisposing to multiple sclerosis. *Hum Mol Genet.* 2015;24(19):5619–27.
35. Heppner FL, Ransohoff RM, Becher B. Immune attack: the role of inflammation in alzheimer disease. *Nat Rev Neurosci.* 2015;16(6):358–72.
36. Hu X, Kim H, Stahl E, Plenge R, Daly M, Raychaudhuri S. Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. *Am J Hum Genet.* 2011;89(4):496–506.
37. Hu X, Kim H, Raj T, Brennan PJ, Trynka G, Teslovich N, Slowikowski K, Chen W-M, Onengut S, Baecher-Allan C, et al. Regulation of gene expression in autoimmune disease loci and the genetic basis of proliferation in CD4+ effector memory T cells. *PLoS Genet.* 2014;10(6):1004404.
38. Mittleman BE, Pott S, Warland S, Zeng T, Mu Z, Kaur M, Gilad Y, Li Y. Alternative polyadenylation mediates genetic regulation of gene expression. *eLife.* 2020;9:57492.
39. Wang X, Goldstein DB. Enhancer domains predict gene pathogenicity and inform gene discovery in complex disease. *Am J Hum Genet.* 2020;106(2):215–33.
40. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020;581(7809):434–43.
41. Calderon D, Nguyen ML, Mezger A, Kathiria A, Müller F, Nguyen V, Lescano N, Wu B, Trombetta J, Ribado JV, et al. Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat Genet.* 2019:1–12.
42. Soskic B, Cano-Gamez E, Smyth DJ, Rowan WC, Nacic N, Esparza-Gordillo J, Bossini-Castillo L, Tough DF, Larmine CG, Bronson PG, et al. Chromatin activity at GWAS loci identifies T cell states driving complex immune diseases. *Nat Genet.* 2019;51(10):1486–93.
43. Vösa U, Claringbould A, Westra H-J, Bonder MJ, Deelen P, Zeng B, Kirsten H, Saha A, Kreuzhuber R, Kasela S, et al. Unraveling the polygenic architecture of complex traits using blood EQTL meta-analysis. *BioRxiv.* 2018;447367.
44. Courtney P, Doherty M. Joint aspiration and injection and synovial fluid analysis. *Best Pract Res Clin Endocrinol Metab.* 2013;27(2):137–69.
45. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods.* 2012;9(4):357.
46. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 2008;9(9):137.
47. Chen M, Zhan Q, Mu Z, Wang L, Zheng Z, Miao J, Zhu P, Li YL. Alignment of single-cell rna-seq samples without overcorrection using kernel density matching. *Genome Res.* 2021;31(4):698–12.
48. Barve RA, Zack MD, Weiss D, Song R-H, Beidler D, Head RD. Transcriptional profiling and pathway analysis of CSF-1 and IL-34 effects on human monocyte differentiation. *Cytokine.* 2013;63(1):10–7.
49. Perrier S, Coussediere C, Dubost J, Albuissou E, Sauvezie B. IL-1 receptor antagonist (IL-1RA) gene polymorphism in Sjogren's syndrome and rheumatoid arthritis. *Clin Immunol Immunopathol.* 1998;87(3):309–13.
50. Lopa S, Leijts MJ, Moretti M, Lubberts E, van Osch G, Bastiaansen-Jenniskens Y. Arthritic and non-arthritic synovial fluids modulate IL10 and IL1RA gene expression in differentially activated primary human monocytes. *Osteoarthritis Cartil.* 2015;23(11):1853–7.
51. Wang G, Sarkar AK, Carbonetto P, Stephens M. A simple new approach to variable selection in regression, with application to genetic fine-mapping. *bioRxiv.* 2019:501114.
52. Foley CN, Staley JR, Breen PG, Sun BB, Kirk PD, Burgess S, Howson JM. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *BioRxiv.* 2019:592238.
53. Tsuchiya H, Ota M, Sumitomo S, Ishigaki K, Suzuki A, Sakata T, Tsuchida Y, Inui H, Hirose J, Kochi Y, et al. Synovial fibroblasts contribute to the genetic risk of rheumatoid arthritis through the synergistic action of cytokines. *bioRxiv.* 2019:861781.
54. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, et al. Next-generation genotype imputation service and methods. *Nat Genet.* 2016;48(10):1284.
55. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ, et al. Plink: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559–75.
56. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38(8):904.
57. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic rna-seq quantification. *Nat Biotechnol.* 2016;34(5):525–7.
58. Frankish A, Diekhans M, Ferreira A-M, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J, et al. Gencode reference annotation for the human and mouse genomes. *Nucleic Acids Res.* 2019;47(D1):766–73.
59. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15–21.
60. Van De Geijn B, McVicker G, Gilad Y, Pritchard JK. Wasp: allele-specific software for robust molecular quantitative trait locus discovery. *Nat Methods.* 2015;12(11):1061.
61. Feng Y-Y, Ramu A, Cotto KC, Skidmore ZL, Kunisaki J, Conrad DF, Lin Y, Chapman W, Uppaulri R, Govindan R, et al. RegTools: Integrated analysis of genomic and transcriptomic data for discovery of splicing variants in cancer. *bioRxiv.* 2018:436634.
62. Delaneau O, Ongen H, Brown AA, Fort A, Panousis NI, Dermizakis ET. A complete tool set for molecular QTL discovery and analysis. *Nat Commun.* 2017;8(1):1–7.
63. Storey JD. False discovery rate. *Int Encycl Stat Sci.* 2011;1:504–8.
64. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, Yoshida S, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature.* 2014;506(7488):376.

65. Loh P-R, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. *Nat Genet.* 2018;50(7):906.
66. Ferreira MA, Vonk JM, Baurecht H, Marenholz I, Tian C, Hoffman JD, Helmer Q, Tillander A, Ullemar V, Van Dongen J, et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat Genet.* 2017;49(12):1752.
67. Paternoster L, Standl M, Waage J, Baurecht H, Hotze M, Strachan DP, Curtin JA, Bønnelykke K, Tian C, Takahashi A, et al. Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. *Nat Genet.* 2015;47(12):1449.
68. Ferreira MA, Mathur R, Vonk JM, Sz wajda A, Brumpton B, Granell R, Brew BK, Ullemar V, Lu Y, Jiang Y, et al. Genetic architectures of childhood-and adult-onset asthma are partly distinct. *Am J Hum Genet.* 2019;104(4):665–84.
69. Demenais F, Margeritte-Jeannin P, Barnes KC, Cookson WO, Altmüller J, Ang W, Barr RG, Beaty TH, Becker AB, Beilby J, et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat Genet.* 2018;50(1):42.
70. Morris DL, Sheng Y, Zhang Y, Wang Y-F, Zhu Z, Tomblason P, Chen L, Graham DSC, Bentham J, Roberts AL, et al. Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus. *Nat Genet.* 2016;48(8):940.
71. Sawcer S, Ban M. Multiple sclerosis genomic map implicates peripheral immune cells & microglia in susceptibility. In: American Association for the Advancement of Science; 2019.
72. Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, Mead D, Bouman H, Riveros-Mckay F, Kostadima MA, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell.* 2016;167(5):1415–29.
73. Watanabe K, Stringer S, Frei O, Mirkov MU, de Leeuw C, Polderman TJ, van der Sluis S, Andreassen OA, Neale BM, Posthuma D. A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet.* 2019;51(9):1339–48.
74. Mahajan A, Wessel J, Willems SM, Zhao W, Robertson NR, Chu AY, Gan W, Kitajima H, Taliun D, Rayner NW, et al. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. *Nat Genet.* 2018;50(4):559–71.
75. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, Sealock J, Karlsson IK, Hägg S, Athanasiu L, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing alzheimer's disease risk. *Nat Genet.* 2019;51(3):404–13.
76. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, Tan M, Kia DA, Noyce AJ, Xue A, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* 2019;18(12):1091–102.
77. Wuttke M, Li Y, Li M, Sieber KB, Feitosa MF, Gorski M, Tin A, Wang L, Chu AY, Hoppmann A, et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet.* 2019;51(6):957.
78. Yengo L, Sidorenko J, Kempner KE, Zheng Z, Wood AR, Weedon MN, Frayling TM, Hirschhorn J, Yang J, Visscher PM, et al. Meta-analysis of genome-wide association studies for height and body mass index in 700,000 individuals of european ancestry. *Hum Mol Genet.* 2018;27(20):3641–9.
79. Guo Q, Schmidt MK, Kraft P, Canisius S, Chen C, Khan S, Tyrer J, Bolla MK, Wang Q, Dennis J, et al. Identification of novel genetic markers of breast cancer survival. *JNCI J Natl Cancer Inst.* 2015;107(5):081.
80. Liu B, Gloude mans MJ, Rao AS, Ingelsson E, Montgomery SB. Abundant associations with gene expression complicate GWAS follow-up. *Nat Genet.* 2019;51(5):768–9.
81. Katz Y, Wang ET, Airoidi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods.* 2010;7(12):1009–15.
82. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 2016;44(W1):160–5.
83. Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, Habermann B, Akhtar A, Manke T. High-resolution tads reveal dna sequences underlying genome organization in flies. *Nat Commun.* 2018;9(1):1–15.
84. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham III CO, Birnbaum NS, Burmester GR, Bykerk VP, Cohen MD, et al. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum.* 2010;62(9):2569–81.
85. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9.
86. Quinlan AR. BEDTools: the Swiss-army tool for genome feature analysis. *Curr Protoc Bioinforma.* 2014;47(1):11–2.
87. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014;30(7):923–30.
88. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 2014;15(2):29.
89. Mu Z, Wei W, Fair BJ, Miao J, Zhu P, Li Y. The impact of cell-type and context-dependent regulatory variants on human immune traits. EGAD00001007003. Datasets. EGA Archive. 2021.

Publisher's Note

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