Limited statistical evidence for shared genetic effects of eQTLs and autoimmune disease-associated loci in three major immune cell types

Sung Chun[1,2,3], **Alexandra Casparino**[4], **Nikolaos A Patsopoulos**[3,5,6], **Damien C Croteau-Chonka**[2,7], **Benjamin A Raby**[2,7], **Philip L De Jager**[2,3,5,6], **Shamil R Sunyaev**[1,2,3,*], and **Chris Cotsapas**[3,4,8,*]

[1]Division of Genetics, Brigham and Women's Hospital, Boston MA, USA

[2]Department of Medicine, Harvard School of Medicine, Boston MA, USA

[3]Broad Institute of Harvard and MIT, Cambridge MA, USA

[4]Department of Neurology, Yale School of Medicine, New Haven CT, USA

[5]Department of Neurology, Brigham and Women's Hospital, Boston MA, USA

[6]Ann Romney Center for Neurological Diseases, Brigham and Women's Hospital, Boston MA, USA

[7]Channing Division of Network Medicine, Brigham and Women's Hospital, Boston MA, USA

[8]Department of Genetics, Yale School of Medicine, New Haven CT, USA

Most autoimmune disease risk effects identified by genome-wide association studies (GWAS) localize to open chromatin with gene regulatory activity. GWAS loci are also enriched for expression quantitative trait loci (eQTLs), suggesting that most risk variants alter gene expression[1,2]. However, because causal variants are difficult to identify and *cis*-eQTLs occur frequently, it remains challenging to identify specific instances of disease-relevant changes to gene regulation. Here, we use a novel joint likelihood framework with

*correspondence to: SRS (ssunyaev@rics.bwh.harvard.edu) and CC (cotsapas@broadinstitute.org).

higher resolution than previous methods to identify loci where autoimmune disease risk and an eQTL are driven by a single, shared genetic effect. Using eQTLs from three major immune subpopulations, we find shared effects in only ~25% of loci. Thus, we uncover a fraction of gene regulatory changes as strong mechanistic hypotheses for disease risk, but conclude that most risk mechanisms likely do not involve changes to basal gene expression.

The autoimmune and inflammatory diseases (AID) are heritable, complex diseases where loss of tolerance to self-antigens results in either systemic or tissue-specific immune attack[3,4]. GWAS have identified hundreds of genomic regions mediating risk to several AID. These associations are primarily non-coding: lead GWAS SNPs are more likely to be associated with expression levels of neighboring genes than expected by chance[12,13], and the same lead SNPs are enriched in regulatory regions marked by chromatin accessibility and modification[1,14]. Fine-mapping reveals enrichment of AID-associated variants in enhancer elements active in stimulated T cell subpopulations[15], with heritability strongly enriched in such regulatory regions[16,17]. Collectively, these strands of evidence suggest that the majority of disease risk is mediated by changes to gene regulation in specific cell subpopulations.

However, these bulk analyses do not formally assess whether expression levels and disease risk can be attributed to a single underlying variant or to independent effects in a locus[18,19]. Though several methods have been developed to assess these alternatives using eQTL data[20–23], they show limited resolution to detect cases where distinct disease and eQTL causal variants are in linkage disequilibrium. Here, we present an approach to test if a GWAS risk association and an eQTL are driven by the same underlying genetic effect, accounting for the LD between causal variants. Using data from ImmunoChip studies of seven AID comprising >180,000 samples in total (Supplementary Table 1), we test if associations in 272 known risk loci are consistent with *cis*-eQTL for genes in each region, measured in three relevant immune cell populations: lymphoblastoid cell lines (LCLs), CD4+ T cells and CD14+ monocytes[24,25].

When associations to two traits – here, disease trait and eQTL – are driven by the same underlying causal variant, the joint evidence of association should be maximized at the markers in tightest LD with the causal variant[19,26]. Here, we directly evaluate this joint likelihood (Supplementary Figure 1), unlike previous approaches that look for similarities in the shape of the association curve over multiple markers[20,21,27,28]. When the underlying causal effect is shared, joint likelihood is maximized when we model the same causal variant in both traits; conversely, when the underlying causal variants are different, we expect maximum joint likelihood when we model their closest proxies. We empirically derive the null distribution of the joint likelihood ratio statistic by comparing disease associations to permuted eQTL data(see Methods, Supplementary Figure 2 and Supplementary Notes). We thus directly evaluate whether two associations in the same locus, observed in different cohorts, are due to the same underlying effect.

To assess the performance of our method, we benchmarked it against three recently reported methods: *coloc*[20], a well-calibrated Bayesian framework that considers spatial similarities in association data across sets of markers; gwas-pw[29], which extends this idea to hierarchical priors and optimizes model parameters; and HEIDI/SMR[22], which applies Mendelian

randomization between traits. We simulated pairs of case-control cohorts with either the same or distinct causal variants driving association, and find that our approach shows the best overall performance (Supplementary Tables 2 and 3). When independent causal variants (i.e. not in LD) drive GWAS and eQTL associations, our own method, *coloc* and gwas-pw all had excellent performance. As the LD between the causal variants increases, our method shows the best performance, maintaining high resolution even when the underlying causal variants are in strong LD (AUC = 0.883 when $0.7 < r^2 < 0.8$, Supplementary Figures 3 and 4), whereas the other methods show substantial false positive rates, reporting distinct effects as shared. We also found that our method is robust to within-continent levels of population structure (Supplementary Figures 5 and 6), and when limiting analysis to a subset of SNPs for computational efficiency (Supplementary Figure 7; *coloc* fares similarly, Supplementary Figure 8). Our method also performs well when multiple independent causal variants affect one or both traits (Supplementary Figures 9–11). In practice, our resolution becomes limited at high LD levels ($r^2 > 0.8$), where the false positive rate increases dramatically. We also have limited resolution when the eQTL effect is very weak ($p > 0.01$, Supplementary Figures 12–15). Thus, within these limits, we can accurately detect cases of shared genetic effects between two traits.

To dissect AID risk loci, we first identified densely genotyped ImmunoChip loci showing genome-wide significant association, excluding the Major Histocompatibility Locus due to the extensive LD structure in the region (immunobase.org; Table 1). We next identified genes in a 1Mb window centered on the most associated variant in each locus. Consistent with previous observations that eQTLs are frequently found in GWAS loci, we found that 260/272 loci had at least one gene with an eQTL (p < 0.01) in at least one cell type, with most such effects common across all three tissues (Table 1). We tested if any eQTLs in these loci appear driven by the same underlying effect as the disease associations. We find evidence for shared effects for only 77/5,749 pairs in 55/260 (21%) loci across all diseases, with the proportion varying from 4/34 (12%) for rheumatoid arthritis loci to 6/10 (60%) for ulcerative colitis loci (false discovery rate < 5%; Tables 1 and 2). Of these 77 shared effects, 45 pass even the more stringent family-wise multiple testing correction (Bonferroni corrected P < 0.05). Thus, our analysis reveals that in the majority of AID loci, variants causally involved in disease phenotypes do not overlap variants responsible for eQTL signals in the three broad cell populations we analyzed, which represent the major arms of the immune lineages. Overall, we find that >75% of tested disease-eQTL pairs appear associated to distinct genetic variants in the same locus (Figure 1).

We sought to explain this lack of overlap between disease associations and eQTLs, despite their frequent co-occurrence in the same loci. In particular, although our method showed good performance in simulated data (Supplementary Figure 4), we remained concerned that this lack of overlap may be due to low statistical power in the eQTL data, which come from cohorts of limited sample size. However, we find that even amongst the most strongly supported eQTLs (nominal $p < 10^{-5}$), <25% show evidence of shared effects with disease associations. Conversely, we find strong evidence for distinct effects for the majority of disease-eQTL pairs, with only a subset of comparisons being ambiguous, suggesting that our method is adequately powered to detect shared effects where they exist (Figure 1a and

Supplementary Figures 16–18). To assess whether power affects the total number of loci, rather than eQTL, that can be resolved, we looked more deeply at our significance threshold settings. We find that more liberal thresholds do not increase the number of true positive results after adjusting for false positive rate, indicating that most loci do not contain *any* gene with an eQTL consistent with the disease association (Figure 1b and Supplementary Figure 19). Cumulatively, our results demonstrate that only a minority of AID risk effects drive eQTLs in the three cell populations we tested, which are drawn from diverse lineages of the immune system.

We next focused on the subset of 77 disease/eQTL pairs in 55 loci where we could detect strong evidence of a shared effect (Table 2). We find that 59/77 (77%) of effects are restricted to one cell population, indicating that tissue-specific eQTLs are important components of the molecular underpinnings of disease (Supplementary Figures 20 and 21). The remaining 18 effects are detected in multiple cell populations; for example, the multiple sclerosis association at rs10783847 on chromosome 12 is consistent with eQTLs for the transcript of methyltransferase-like 21B (*METTL21B*) in both CD4[+] T cells and CD14[+] monocytes, but not for the remaining 31 genes in the immediate locus (Figure 2). Although *METTL21B* is expressed in LCLs, there is no evidence of an eQTL in this tissue within 1Mb from rs10783847. Similarly, for the multiple sclerosis association at rs1966115 on chromosome 8 and eQTLs for *ZC2HC1A*, and for the inflammatory bowel disease association at rs55770741 on chromosome 5 and eQTLs for *ERAP2*, we detect a shared effect in all three cell populations. In several cases we find tissue-specific shared effects despite strong eQTLs for the same gene in other tissues: for *ZFP90* and ulcerative colitis risk at rs889561 on chromosome 16, we also find shared effects in CD4[+] and CD14[+] but not LCLs, where we observe a *ZFP90* eQTL at p = 0.005 that has a low likelihood of shared effect with GWAS (joint likelihood P = 0.85). Instead, we find evidence of sharing between disease risk and an eQTL for *NFAT5* in LCLs. Thus, despite the presence of eQTLs for a gene in multiple tissues, not all these effects are consistent with disease associations suggesting that disease-relevant eQTLs are tissue specific.

We also find cases where an eQTL is consistent with associations to multiple diseases. The ankyrin repeat domain 55 (*ANKRD55*) transcript encoded on chromosome 5 has an eQTL in CD4[+] T cells that is shared with associations to multiple sclerosis, Crohn disease and rheumatoid arthritis (Figure 3, all observations are significant after Bonferroni correction). We also find weaker evidence for shared effects between all three diseases and an eQTL for interleukin 6 signal transducer (*IL6ST*) in CD4[+] T cells, which passes the false discovery rate threshold but not Bonferroni correction (Supplementary Figure 22). Similarly, a CD4[+] eQTL for *ELMO1* on chromosome 7 is consistent with associations to both celiac disease and multiple sclerosis (Supplementary Figure 23), a CD14[+] eQTL for *RGS1* on chromosome 1 is consistent with associations to both celiac disease and multiple sclerosis (Supplementary Figure 24), and three other eQTLs are consistent with associations in multiple diseases (Supplementary Figures 25–27). In all cases, these are the only genome-wide significant disease associations reported in these loci. As we consider each disease association independently, these results indicate that the same underlying risk variants drive risk to multiple diseases in these loci by altering gene expression, consistent with observations of shared effects across diseases[7].

Overall, our results suggest that some autoimmune and inflammatory disease loci are consistent with eQTLs acting in specific immune cell subpopulations, which form strong mechanistic hypotheses for the molecular mechanisms driving disease risk. However, these only account for a small fraction of eQTLs present in disease risk loci; this suggests that abundant caution must be exercised before inferring pathological relevance for an observed eQTL simply due to proximity to a disease association. Strong evidence of a shared genetic effect should therefore be established prior to embarking on time-consuming and costly experimental dissection of such effects.

Previous efforts to detect shared effects between traits in specific loci rely on conditional analyses[31] or indirectly leverage linkage disequilibrium to test if the shape of association peaks in the region are similar[20,27,28,32]. In contrast, we directly evaluate whether the data support a shared effect through joint likelihood estimation. Through this direct evaluation, we can resolve cases where two associations are proximal with higher resolution (Supplementary Figures 3 and 4, Supplementary Tables 2 and 3). As our method is general, it may be useful in other contexts, such as establishing if the shared heritability between diseases is driven by the same underlying causal effects[33].

More broadly, our results raise the question of how causal disease variants alter cell function to induce risk, given the strong enrichment of disease risk signal in gene regulatory regions[1], and gene enhancers in particular[15]. We suggest that although gene regulatory regions harboring risk variants are accessible in multiple immune cell subpopulations, they may control gene expression in either a tissue-specific or condition-specific manner. These gene regulatory events may be restricted to very specific cell populations, and easily accessible subsets – such as those we have analyzed here – may not adequately capture these events. Our results therefore reinforce the view that we must seek the appropriate cell type and physiological conditions in order to capture the pathologically relevant gene regulatory changes driving disease risk.

Online Methods

Simulated dataset

We randomly sampled 97 genomic loci of length 200kb across the genome to base our simulations. We excluded sub-telomeric/centromeric regions, sex chromosomes, and regions of sparse genetic map coverage. In each locus, we simulated disease (20,000 cases and 20,000 controls each) and eQTL (250 individuals each) cohorts using HapGen2 [34] and phased haplotypes from the CEU population ($2n$=198) of the 1,000 Genomes Project[35]. For disease cohorts, we set the variant nearest the center of the interval as causal with an odds ratio of 1.1 for each minor allele copy, and simulated five replicate cohorts of cases and controls in each locus. In each locus, we then simulated three different genetic models for the eQTL cohort: no causal variant ("H_0"), the same causal variant as disease ("H_1"), and distinct causal variants between disease and eQTL ("H_2"). For H_2, we selected eQTL causal variants within 50kb of the disease-causing variant and with differing levels of LD between the two causal variants (r^2 of 0 – 0.4, 0.4 – 0.5, 0.5 – 0.6, 0.6 – 0.7, 0.7 – 0.8, or 0.8 – 0.9 in CEU). We selected all disease and eQTL causal variants to have minor allele frequency (MAF) > 10% in CEU. We generate genotypes for the eQTL cohorts using HapGen2 with a

null effect size, then simulate a quantitative phenotype using the allelic mean difference model implemented in GCTA[36] with effect sizes of 0.05, 0.1, or 0.2 in cis-heritability (h^2). In each locus, we generated five replicate eQTL cohorts for H_0 and H_1 each; for H_2, we generated a single cohort up to five distinct causal variants per locus.

We used plink to calculate the genetic association with disease and expression phenotypes in logistic and linear regression models, respectively, after filtering out SNPs with MAF < 5% in each cohort. We rejected cohorts showing weak maximum association signals (association $p > 10^{-5}$ for disease cohorts and association $p > 0.01$ for eQTL cohorts). In addition, as expected from the coalescent forward simulation model on which HapGen2 is based, a fraction of our simulated cohorts showed maximal association to a SNP in low LD with the causal variant we had specified ($r^2 < 0.8$ measured in-sample). We kept these cohorts as gene expression traits only, to better capture the vagaries of resolution limits inherent in the small sample size of eQTL studies but excluded the disease cohorts. Overall, we rejected 20% of disease cohorts and 11% of eQTL cohorts and generated a total of 5,680, 829, and 4,666 disease-eQTL comparisons under H_0, H_1, and H_2, respectively.

To test the effect of mild population mismatch, we also generated a second set of eQTL cohorts, this time using base haplotypes of all non-Finnish Europeans from the 1000 Genomes Project (CEU+GBR+TSI+IBS, $2n = 808$).

To explore the scenarios that multiple independent causal variants in a locus affect the same phenotype, we generated another sets of simulated disease-eQTL pairs assuming two causal variants for disease, eQTL, or both traits in the genetic background of 48 genomic loci. We set the disease-causing variant at the center of locus as the reference SNP and added a second causal variant, varying the LD with the reference SNP (r^2 bins of 0 – 0.4, 0.4 – 0.5, 0.5 – 0.6, 0.6 – 0.7, 0.7 – 0.8, or 0.8 – 0.9). For disease cohorts, we set the OR of the central risk variant to 1.1, as in the original simulations, and the OR of the second causal variant to 1.1, 1.049, or 1.024 (i.e. 1x, 0.5x, or 0.1× the effect size of the central variant on a log scale). For expression phenotypes, the total h^2 of 0.1 was split between two independent causal variants. The relative effect size of second causal variant was scaled to 1x, 0.5x, or 0.1× relative to the central causal SNP, without standardizing the genotypes. For each combination of causal variants, we generated two replicate cohorts for disease and a single cohort for eQTL. Again, we rejected cohorts where we observe the strongest association at a variant in low LD with the specified causal variants ($r^2 < 0.8$).

Disease GWAS dataset

We downloaded association summary statistics for type 1 diabetes (T1D), rheumatoid arthritis (RA), celiac disease (CEL), multiple sclerosis (MS), inflammatory bowel disease (IBD), Crohn's disease (Crohn), and ulcerative colitis (UC) from ImmunoBase (immunobase.org; Supplementary Table 1). For MS, we used the association statistics derived from the combined cohort of discovery and validation samples[8] in order to maximize the sample size and genetic resolution. For IBD, Crohn, and UC, summary data are from European subset of the latest trans-ethnic association study[30]. All association data are solely based on ImmunoChip samples and do not include imputed genotypes. To address population structure, we limited our analyses to European subjects only with the exception

of RA, which includes 620 Punjab individuals out of a total of 27,345. T1D summary statistics are from the meta-analysis between case/control association and affected sib-pair analysis.

As our method works best on dense genotype data, we restricted our analyses to the 188 loci genotyped at high density on ImmunoChip. We excluded the Major Histocompatibility Complex (MHC) locus, due to the complex landscape of selection and resulting complex LD patterns. For each disease, we sought the largest published genetic mapping study and identified genome-wide significant associations reported in the 188 ImmunoChip loci. We note that these reports may contain additional samples, so the associations may not be genome-wide significant in the ImmunoChip studies alone. We also excluded any secondary associations after conditioning on initial results, as these are inconsistently reported across diseases. If multiple independent associations are reported within the same ImmunoChip region for any disease, we divide the region at the mid-point between the reported markers and select lead SNPs in each sub-interval separately.

eQTL dataset

We examined eQTLs in Lymphoblastoid Cell Lines (LCLs) and primary CD4[+] T cells and CD14[+] monocytes obtained from healthy donors[24,25] (Supplementary Table 1). For LCLs, we obtained imputed genotypes and normalized RNAseq in RPKM for 278 non-Finnish European donors in the Geuvadis project. We removed SNPs with minor allele frequency (< 5%), high probability of Hardy-Weinberg disequilibrium ($P_{HW} < 10^{-5}$), or high genotype missing rate (>5%). We removed pseudogenes and transcripts without assigned gene symbols from the expression data, and calculated association statistics by linear regression of genotype on expression levels, including three population principal components to control for structure[37,38]. For CD4[+] and CD14[+], we regressed normalized expression levels for European Americans (n=213 and 211, respectively) on similarly QCed imputed allele dosages. For all cell types, we generated adaptive permutation statistics from 10^3 up to 10^6 iterations, using all covariates[37].

Joint likelihood mapping (JLIM)

To test the hypothesis that association signals for two traits are driven by the same causal variant, we contrasted the joint likelihood of observed association statistics under the assumption of same compared to distinct causal variant. Due to limited genetic resolution, distinct causal variants were defined by separation in LD space by $r^2 < \theta$ from each other. The limit of genetic resolution θ is a user-specified parameter and was set to 0.8 in this study. We assumed that at most one causal variant was present in the locus for each trait and that samples of two trait association were not overlapping. We designed the joint likelihood mapping (JLIM) statistic Λ in an asymmetrical fashion, requiring only summary-level statistics for one trait (primary trait) but genotype-level data for the other (secondary trait). Specifically, Λ was defined as the sum of log likelihood that the causal variant underlying secondary trait is more likely to be same as than distinct from the variant underlying primary trait, as integrated over a set of likely causal variants under a GWAS peak of primary trait:

$$\Lambda = \sum_{i \in \mathcal{N}_\theta^1(m^*)} L_1(i) \cdot \log \frac{L_1(i) L_2(i)}{\max_{j \notin \mathcal{N}_\theta^2(i)} L_1(i) L_2(j)}$$

$$= \sum_{i \in \mathcal{N}_\theta^1(m^*)} L_1(i) \cdot (\log L_2(i) - \max_{j \notin \mathcal{N}_\theta^2(i)} \{\log L_2(j)\})$$

where m^* is the most associated SNP for primary trait, $L_1(i)$ and $L_2(i)$ are the likelihood of SNP i being causally associated with primary and secondary traits, respectively, and $\mathcal{N}_\theta^1(i)$ and $\mathcal{N}_\theta^2(i)$ are the sets of SNPs within LD neighborhood around SNP i, as defined by $SNP\, j | r_{i,j}^2 > \theta$. We derived $\mathcal{N}_\theta^1$ from the reference LD panel and $\mathcal{N}_\theta^2$ directly from the genotypes of secondary trait cohort. We used disease outcome as primary trait, leveraging the larger sample size and dense genotyping, and gene expression as secondary trait, taking advantage of the availability of individual genotype data.

The likelihood of causal association was calculated by approximating the local LD structure with pairwise correlation similarly as Kichaev et al. and Hormozdiari et al. Briefly, when SNP c is the only causal variant in the locus with non-centrality λ_c, association static z_i of non-causal SNP i follows a normal distribution $N(r_{i,c} \lambda_c, 1)$, where $r_{i,c}$ is LD between SNPs i and c measured in pairwise Pearson correlation of genotypes. In general, when association statistics $\boldsymbol{Z} = (z_1, z_2, \ldots z_M)^T$ are provided for all M SNPs in the analysis window, the likelihood of SNP i being the causal variant with non-centrality λ_i is:

$$L(\boldsymbol{Z}; \lambda_i \neq 0) = \phi_{MVN}(\boldsymbol{Z}; mean = \sum(\lambda_i \circ C), var = \sum)$$

where ϕ_{MVN} is the multivariate normal density function, C is an incident vector with $C_k = 1$ if and only if $K = I$, Σ is a $M \times M$ local LD matrix defined by pairwise Pearson correlation between genotypes, and $\circ$ is element-wise multiplication[39]. Since we do not know the true non-centrality of causal variant, we estimated the profile likelihood, which simplifies to a closed form[40]:

$$\log L(\boldsymbol{Z}; \lambda_i^{mle}) = \frac{1}{2}(-\boldsymbol{Z}^T \sum{}^{-1} \boldsymbol{Z} + z_i^2) - \frac{1}{2}\log((2\pi)^M |\sum|)$$

with $\lambda_i^{mle} = z_i$. Thus, given association statistics for primary and secondary traits, $\boldsymbol{Z} = (z_1, z_2, \ldots z_M)^T$ and $\boldsymbol{W} = (w_1, w_2, \ldots w_M)^T$, the test statistic Λ simplifies to:

$$\Lambda = \sum_{i \in \mathcal{N}_\theta^1(m^*)} e^{\frac{1}{2}(z_i^2 - z_{m^*}^2)} \cdot (w_i^2 - \max_{j \notin \mathcal{N}_\theta^2(i)} w_j^2)$$

The p-value of joint likelihood is estimated by permuting phenotypes of secondary traits as under the trivial null hypothesis that that there is no casual variant for secondary trait in the locus (H_0). With respect to the more likely null that distinct causal variants underlie association signals of two traits (H_2), we can show that asymptotically as the non-centrality

of causal variant increases, p-values estimated from H_0 behave conservatively with respect to H_2 (Supplementary Notes):

$$P_{JLIM} = P(\Lambda \geq l|H_0) \geq P(\Lambda \geq l|H_2)$$

Thus, with large enough sample or effect sizes, joint likelihood test against H_0 will also reject H_2 in favor of alternative hypothesis of shared causal variant (H_1). Further, to evaluate whether this property holds for practical non-centrality values, we examined our negative controls simulating H_2, specifically, if P_{JLIM} was highly shifted toward 1.0 (Supplementary Figure 2) and similar or larger than empirically estimated false positive rates as expected (Supplementary Table 3   0.05).

For both simulated and real GWAS data, we applied JLIM to SNPs with data for both primary and secondary traits, present in the reference LD panels, and within 100kb of the most associated marker to disease ("lead SNP"). In ImmunoChip data, the analysis windows were further confined by the boundaries of the dense genotyping intervals. We compared each lead SNP to eQTL data for all genes with a transcription start sites (TSS) up to 1Mb from the lead SNP, and an eQTL association $p < 0.01$ for at least one SNP in the analysis window. To minimize computational burden, we did not consider SNPs associated with neither disease or eQTL (association $p > 0.1$ to both). For the reference LD panel, we used the base haplotypes of HapGen simulation for simulated datasets, and non-Finnish European samples (n=404) of the 1000 Genomes Project (phase 3, release 2013/05/02) for ImmunoChip loci.

We corrected for multiple tests using false discovery rate (FDR) levels and Bonferroni correction. The FDR was calculated separately for specific disease and cell type combination as:

$$FDR(p) = \frac{p \, N}{\#\{P_{JLIM} \leq p\}}$$

where p is a JLIM p-value cut-off, and N is the number of all tested disease lead SNP-eQTL candidate gene combinations. The FDR was calculated for each cell type since the distribution of JLIM p-values can vary depending on the disease relevance of cell type. To provide a list of higher confidence hits in each disease, we also applied the Bonferroni correction to nominal JLIM p-values for the number of tests across all three cell types.

Benchmark comparison

We used our simulations to compare the performance of our method (here abbreviated as JLIM, for joint likelihood mapping) to three existing methods: Bayesian *coloc*, gwas-pw, and SMR/HEIDI. We ran *coloc* (version 2.3-1) using default parameter settings with the colocalization prior p_{12} set to 10^{-6}. We followed the authors' recommendation to use beta and variance of beta for the case/control cohorts as summary statistics. For quantitative trait cohorts, we also provided in-sample minor allele frequencies. We applied gwas-pw (version 0.21) with default parameters. All simulated disease-eQTL pairs were combined into a

single batch and analyzed together so that gwas-pw can optimize the model parameters. We ran SMR/HEIDI (version 0.64) with default parameters except the p-value threshold to select the top associated eQTL (peqtl-smr, default 5×10^{-8}), which was relaxed to 0.01 in order to enable the test on simulated disease-eQTL pairs with weak eQTL association. Tests producing significant heterogeneity by HEIDI ($p_{HEIDI} < 0.05$) were called negative regardless of p_{SMR} values since they are likely to harbor distinct causal variants between disease and eQTL. For *coloc* and gwas-pw, predictions were made based only on reported posterior probability of colocalization (PP4 and PP3, respectively) although they report posteriors for other competing models. For overall performance comparison, we evaluated the area under the receiver operator curve (ROC; Supplementary Figures 3, *p* and 4) using H_1 as known positives and H_0 and H_2 as known negatives. For SMR/HEIDI, the sensitivity did not reach 1.0 even at the specificity of 0.0 since the method called significant heterogeneity on 15% of H_1 ($p_{HEIDI} < 0.05$). The sensitivity and specificity were also compared at *p*-value cut-offs of 0.01 and 0.05 (Supplementary Tables 2 and 3). For *coloc* and gwas-pw, the posterior probability cut-offs equivalent to *p*-values were determined from the false positive rates on null simulation of no eQTL.

Bayesian *coloc* on real data

As ImmunoChip data is only available as summary statistics, we used ran *coloc*[20] with the p-values of association the minor allele frequencies from non-Finnish Europeans from the 1000 Genomes Projectfor disease cohorts, and with quantitative beta and variance of beta calculated onfor eQTL association datacohorts. We also provided to coloc the minor allele frequencies of non-Finnish Europeans from the 1,000 Genomes Project[35]. The and a colocalization prior p_{12} was set to= 10^{-6} [6], and the prediction was made at PP4 0.75 for higher confidence (Supplementary Table 4). We did not consider the type 1 diabetes data, where case/control sample size is limited after excluding affected sib pair data.

Estimating the number of disease GWAS loci with consistent eQTL effects

We expect JLIM p-values to follow a bimodal distribution with modes close to zero and one when the data support a model of shared or distinct causal effects, respectively. Conversely, under the null model of no *cis*-eQTL association, we expect a uniform *p*-value distribution. We can thus estimate the proportion of disease-eQTL pairs belonging to the null π_0, same π_1 and distinct π_2 causal variant models from the observed *p*-value distribution[41] (Supplementary Figures 16–18). To assess if the strength of the eQTL association influences the likelihood of identifying a shared causal variant, we calculate these proportions for subsets of trait pairs defined by minimum eQTL *p*-value. In each bin, we identified the limits of the uniform portion of the distribution γ_1 and γ_2 and estimate π_0, π_1 and π_2 as:

$$\pi_0 = \frac{\#\{\gamma_1 < P_{JLIM} < \gamma_2\}}{(\gamma_2 - \gamma_1)N}$$
$$\pi_1 = \frac{\#\{P_{JLIM} \leq \gamma_1\}}{N} - \gamma_1 \pi_0$$
$$\pi_2 = \frac{\#\{P_{JLIM} \geq \gamma_2\}}{N} - (1 - \gamma_2)\pi_0$$

To estimate the number of disease GWAS loci that can be explained by consistent effect of same causal variant on disease and eQTL (denoted by $\mathscr{C}$ below), we incrementally relaxed the p-value cut-offs of JLIM and examined the trends of the number of disease loci with at least one JLIM hit and subtracted the expected number of false positive loci (Figure 1 and Supplementary Figure 19). Specifically, at each JLIM p-value cutoff p_i, we successively calculated $\mathscr{C}(p_i)$:

$$\mathscr{C}(p_i)=\mathscr{C}(p_{i-1})+|\mathscr{D}(p_i)\cap\mathscr{D}(p_{i-1})^c|-\sum_{d\in\mathscr{D}(p_{i-1})^c}\mathscr{E}(d,p_i)+\sum_{d\in\mathscr{D}(p_{i-1})^c}\mathscr{E}(d,p_i)$$

where $p_{i-1} < p_i$ with $p_0 = 0$, $\mathscr{D}(p)$ is the set of disease GWAS loci with at least one eQTL gene in any cell type passing the JLIM p-value cut-off p, and $\varepsilon(d,p)$ is the probability that disease GWAS locus d has a false positive eQTL gene passing the JLIM p-value cutoff p. We estimated the lower and upper bounds of $\varepsilon(d,p)$ using the Monte Carlo method by randomly selecting false positive eQTL genes within the locus d at rates of $(1 - \pi_1) \cdot lb$ or $(1 - \pi_1) \cdot ub$ over 1,000 iterations. The lb and ub are the lower and upper bounds of false positive rate of JLIM against true null. Note that π_1 and lb depend on the cell type and strength of eQTL association.

As the true null is mixture of two nulls, H_0 and H_2, the false positive rate of JLIM against true null $P(\Lambda \geq l|H_0 \cup H_2)$ can be bounded by using the following decomposition:

$$P(\Lambda \geq l|H_0\cup H_2)=P(\Lambda \geq l|H_0)\frac{P(H_0)}{P(H_0)+P(H_2)}+P(\Lambda \geq l|H_2)\frac{P(H_2)}{P(H_0)+P(H_2)}$$

While the false positive rate under distinct null $P(\Lambda \geq l|H_2)$ is difficult to estimate, it is non-negative by definition and asymptotically bounded by permutation p-value $P(\Lambda \geq l|H_0)$, i.e. P_{JLIM}, as the non-centrality of causal variant increases. Therefore, we took:

$$ub=P_{JLIM}$$
$$lb=P_{JLIM}\frac{\pi_0}{\pi_0+\pi_2}=P_{JLIM}\frac{\pi_0}{1-\pi_1}$$

and estimated the bounds of locus-level false positive rates $\varepsilon(d,p)$ and number of disease loci with consistent effects $\mathscr{C}(p_i)$.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Bibliography

1. Maurano MT, et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science. 2012; 337:1190–1195. [PubMed: 22955828]

2. Bernstein BE, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol. 2010; 28:1045–1048. [PubMed: 20944595]

3. Walsh SJ, Rau LM. Autoimmune diseases: a leading cause of death among young and middle-aged women in the United States. Am J Public Health. 2000; 90:1463–1466. [PubMed: 10983209]

4. Autoimmune Diseases Coordinating Committee. Report of the Autoimmune Diseases Coordinating Committee. American Autoimmune Related Diseases Association (AARDA) & National Coalition of Autoimmune Patient Groups (NCAPG); 2011.

5. Eaton WW, Rose NR, Kalaydjian A, Pedersen MG, Mortensen PB. Epidemiology of autoimmune diseases in Denmark. Journal of Autoimmunity. 2007; 29:1–9. [PubMed: 17582741]

6. Criswell LA, et al. Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. AJHG. 2005; 76:561–571.

7. Cotsapas C, et al. Pervasive Sharing of Genetic Effects in Autoimmune Disease. 2011; 7:e1002254.

8. International Multiple Sclerosis Genetics Consortium (IMSGC) et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nature Genetics. 2013; 45:1353–1360. [PubMed: 24076602]

9. Fortune MD, et al. Statistical colocalization of genetic risk variants for related autoimmune diseases in the context of common controls. Nature Genetics. 2015; doi: 10.1038/ng.3330

10. Cotsapas C, Hafler DA. Immune-mediated disease genetics: the shared basis of pathogenesis. Trends Immunol. 2013; 34:22–26. [PubMed: 23031829]

11. Cortes A, Brown MA. Promise and pitfalls of the Immunochip. Arthritis Res Ther. 2011; 13:101. [PubMed: 21345260]

12. Trynka G, et al. Disentangling the Effects of Colocalizing Genomic Annotations to Functionally Prioritize Non-coding Variants within Complex-Trait Loci. Am J Hum Genet. 2015; 97:139–152. [PubMed: 26140449]

13. Nicolae DL, et al. Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. PLoS Genetics. 2010; 6:e1000888. [PubMed: 20369019]

14. Trynka G, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. Nature Genetics. 2013; 45:124–130. [PubMed: 23263488]

15. Farh KKH, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2015; 518:337–343. [PubMed: 25363779]

16. Gusev A, et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. Am J Hum Genet. 2014; 95:535–552. [PubMed: 25439723]

17. Finucane HK, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nature Genetics. 2015; 47:1228–1235. [PubMed: 26414678]

18. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. Nature Reviews Genetics. 2013; 14:483–495.

19. Nica AC, Dermitzakis ET. Using gene expression to investigate the genetic basis of complex disorders. Hum Mol Genet. 2008; 17:R129–34. [PubMed: 18852201]

20. Giambartolomei C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genetics. 2014; 10:e1004383. [PubMed: 24830394]

21. Guo H, et al. Integration of disease association and eQTL data using a Bayesian colocalisation approach highlights six candidate causal genes in immune-mediated diseases. Hum Mol Genet. 2015; 24:3305–3313. [PubMed: 25743184]

22. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nature Genetics. 2016

23. He X, et al. Sherlock: detecting gene-disease associations by matching patterns of expression QTL and GWAS. Am J Hum Genet. 2013; 92:667–680. [PubMed: 23643380]

24. Lappalainen T, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature. 2013; 501:506–511. [PubMed: 24037378]

25. Raj T, et al. Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. Science. 2014; 344:519–523. [PubMed: 24786080]

26. Nica AC, et al. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. PLoS Genetics. 2011; 7:e1002003. [PubMed: 21304890]

27. Malik R, et al. Shared genetic basis for migraine and ischemic stroke: A genome-wide analysis of common variants. Neurology. 2015; 84:2132–2145. [PubMed: 25934857]

28. Winsvold BS, et al. Genetic analysis for a shared biological basis between migraine and coronary artery disease. Neurol Genet. 2015; 1:e10. [PubMed: 27066539]

29. Pickrell JK, et al. Detection and interpretation of shared genetic influences on 42 human traits. Nature Genetics. 2016; 48:709–717. [PubMed: 27182965]

30. Jostins L, et al. Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012; 491:119–124. [PubMed: 23128233]

31. Nica AC, et al. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. PLoS Genetics. 2010; 6:e1000895. [PubMed: 20369022]

32. Wallace C, et al. Statistical colocalization of monocyte gene expression and genetic risk variants for type 1 diabetes. Hum Mol Genet. 2012; 21:2815–2824. [PubMed: 22403184]

33. Bulik-Sullivan B, et al. An atlas of genetic correlations across human diseases and traits. Nature Genetics. 2015; 47:1236–1241. [PubMed: 26414676]

34. Su Z, Marchini J, Donnelly P. HAPGEN2: simulation of multiple disease SNPs. Bioinformatics. 2011; 27:2304–2305. [PubMed: 21653516]

35. Consortium T1GP et al. A map of human genome variation from population-scale sequencing. Nature. 2010; 467:1061–1073. [PubMed: 20981092]

36. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A Tool for Genome-wide Complex Trait Analysis. The American Journal of Human Genetics. 2011; 88:76–82. [PubMed: 21167468]

37. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. AJHG. 2007; 81:559–575.

38. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics. 2006; 38:904–909. [PubMed: 16862161]

39. Kichaev G, et al. Integrating Functional Data to Prioritize Causal Variants in Statistical Fine-Mapping Studies. PLoS Genetics. 2014; 10:e1004722. [PubMed: 25357204]

40. Chen X, et al. Dominant Genetic Variation and Missing Heritability for Human Complex Traits: Insights from Twin versus Genome-wide Common SNP Models. 2015; 97:708–714.

41. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences. 2003; 100:9440–9445.
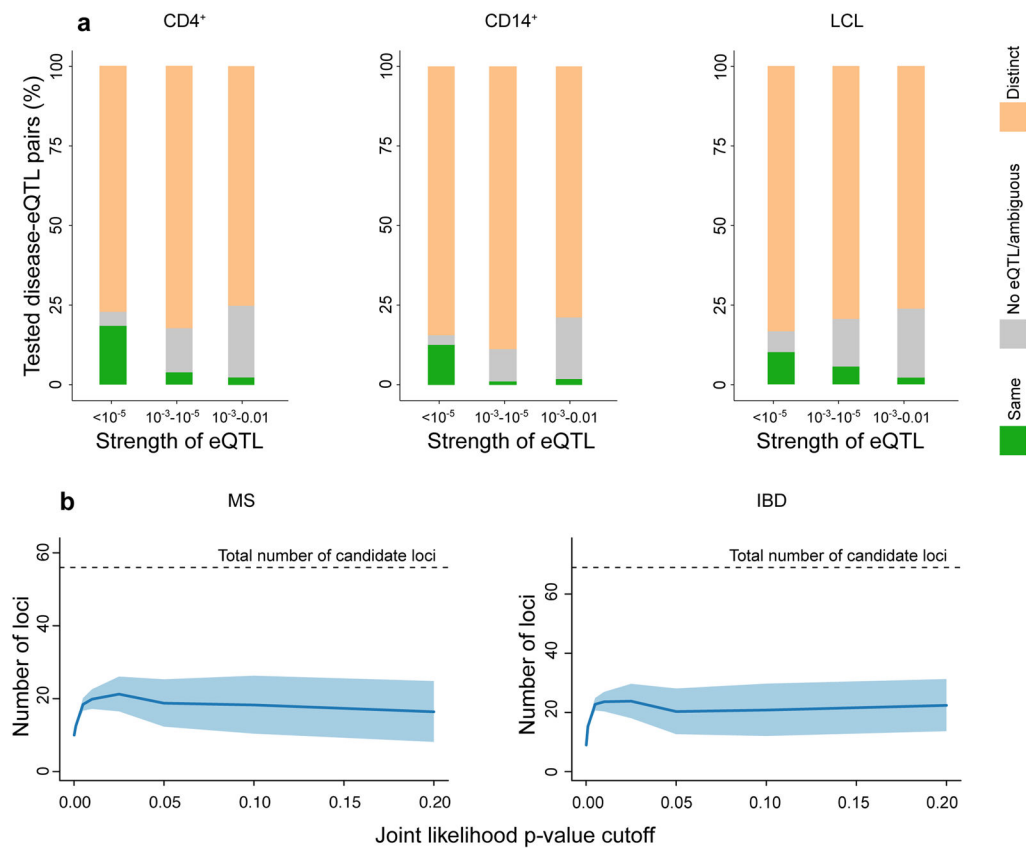
Figure 1. Only a minority of disease associations share genetic effects with eQTLs across three immune cell subpopulations

(**a**) We find strong evidence that approximately 75% of eQTLs are driven by distinct genetic effects (orange) to 260 disease risk associations across 154 ImmunoChip regions. The proportion of shared effects (green) we are able to detect is less than 25%, even for relatively strong eQTLs with nominal association $p < 10^{-5}$. We find no compelling evidence for either shared or distinct associations for a small proportion of disease-eQTL pairs (gray). (**b**) The median number of loci with at least one shared effect eQTL in any cell type (blue line) at more liberal significance thresholds remains constant after false positive adjustment, further supporting this conclusion. The shaded area represents the lower and upper expectation bounds for disease-eQTL pairs driven by the same causal variant. Only 31–47% of multiple sclerosis associations and 30–45% of inflammatory bowel disease associations are consistent with eQTL effects. Equivalent data for the other diseases are presented in Supplementary Figure 19.
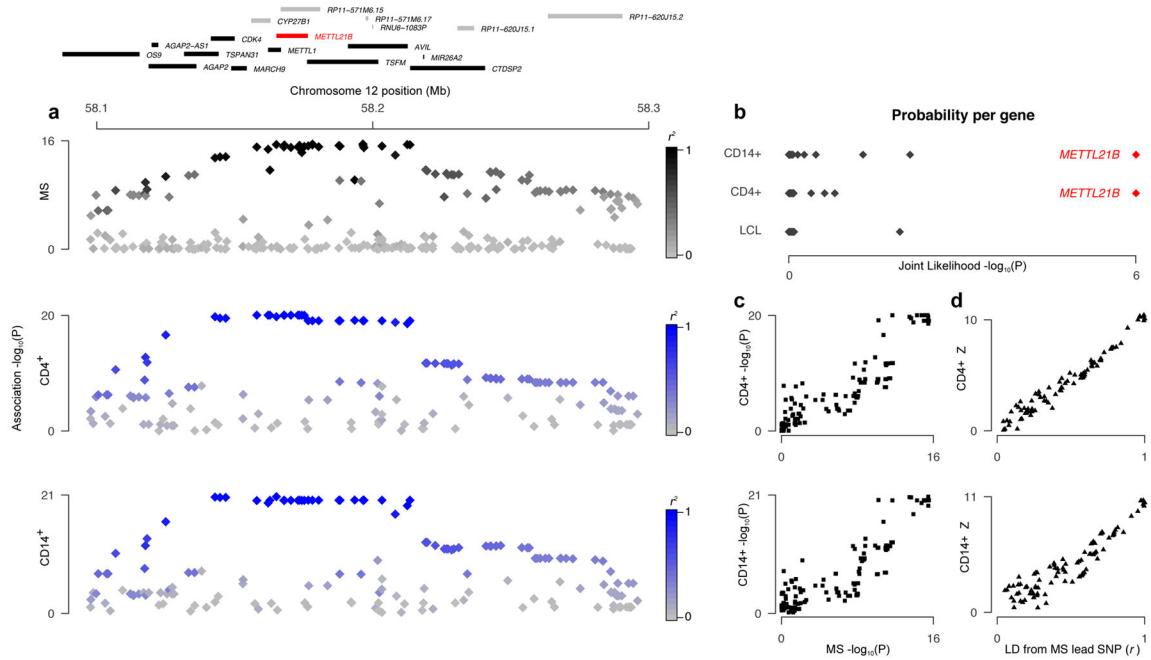
Figure 2. A multiple sclerosis association on chromosome 12 is consistent with eQTLs for METTL21B in both CD4+ T cells and CD14+ monocytes

(a) A genome-wide significant association to multiple sclerosis risk (upper panel; shading denotes strength of LD to the most associated variant rs10783847). This association is consistent with eQTLs for *METTL21B* in CD4+ T cells (middle panel) and CD14+ monocytes (lower panel, both shaded by LD to rs10783847), but not to eQTL data for any other genes in the region (upper gene track: black boxes denote 31 genes with eQTL data available in addition to *METTL21B* (red); gray denotes genes which are not reliably detected in our data or do not have eQTL $p < 0.01$ in the region). **(b)** Joint likelihood *p*-values for 32 candidate genes analyzed for this MS association peak in three cell types. Those with FDR < 5% are shown in red. **(c)** Association *p*-values for MS risk (x-axis) and eQTLs (y-axis) are strongly correlated for both CD4+ T cells (middle panel) and CD14+ monocytes (lower panel). **(d)** Similarly, eQTL association *Z* statistics scale linearly with LD (*r*, × axis) to rs10783847, consistent with a model of a single causal variant driving both disease association and eQTL.
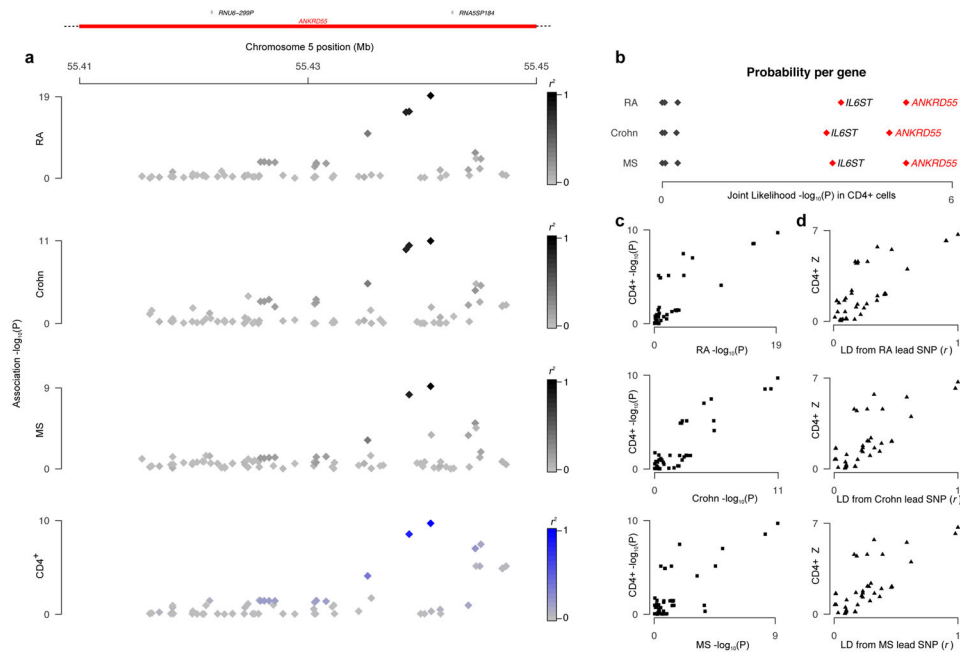
Figure 3. Associations to multiple sclerosis, Crohn disease and rheumatoid arthritis (RA) on chromosome 5 are consistent with an eQTL for *ANKRD55* in CD4$^+$ T cells

(a) Genome-wide significant associations to all three diseases (upper panels) and eQTL data for *ANKRD55* (lower panel; shading in all panels proportional to LD to the most associated variant rs71624119). Due to the variable density of ImmunoChip data, the analysis window is small and only overlaps the coding region of *ANKRD55*, though we test eQTLs for five genes with a transcriptional start site within 1Mb of the the association. **(b)** Joint likelihood *p*-values for five candidate genes analyzed for this locus in CD4$^+$ T cells. Those with FDR < 5% are shown in red. **(c)** Association *p*-values for each disease (x axis) are strongly correlated to those for the *ANKRD55* eQTL in CD4$^+$ cells (y axis). **(d)** Similarly, eQTL association *Z* statistics scale linearly with LD (*r*, × axis) to rs71624119 for all three diseases, consistent with a model of a single causal variant driving all disease associations and the eQTL.

Table 1

Only a minority of disease associations share causal variants with eQTLs across three immune cell subpopulations

We identified 260 disease associations in ImmunoChip regions with at least one eQTL within 100kb of the most associated SNP. Only 55/260 (21%) of these associations show evidence of a shared effect with an eQTL in that region. Thus, while eQTLs are abundant in disease-associated loci, they do not appear to be driven by the same causal variant as the disease association.

| | | Number of loci | | | | | | | |
| | | eQTL present[2] | | | | Driven by same effect | | | |
Disease	Densely genotyped[1]	CD4+	CD14+	LCL	Total	CD4+	CD14+	LCL	Total
MS	59	52	54	53	56	11	3	6	14
IBD	69	68	66	66	69	10	9	1	15
Crohn	19	18	18	17	18	2	1	0	3
UC	10	9	9	10	10	4	1	3	6
T1D	47	39	39	35	40	2	0	5	7
RA	34	34	34	34	34	3	0	1	4
CEL	34	33	32	33	33	5	2	0	6
Overall	272	253	252	248	260	37	16	16	55

[1] We only consider associations reported at genome-wide significant levels and overlapping genomic regions densely genotyped on ImmunoChip, excluding conditional peaks and MHC loci (see Methods).

[2] eQTLs are selected if there is nominal association (eQTL $p < 0.01$) to at least one SNP within 100kb of the most associated SNP to disease, and a transcription start site of the gene within 1Mb of that SNP.

[3] Number of loci where disease association is consistent with a shared effect for at least one eQTL (FDR < 5%). Associations common to Crohn disease and ulcerative colitis are listed under inflammatory bowel disease[30].

Table 2

Fifty five loci harbor eQTLs driven by the same variants as an association to at least one of seven diseases

We find 77 instances of shared disease-eQTL effects in 55 loci (joint likelihood of shared association FDR < 5%).

Disease	Lead SNP[1]	Gene	CD4+ T cell		CD14+ monocytes		LCL	
			eQTL p[2]	JLIM p[3]	eQTL p[2]	JLIM p[3]	eQTL p[2]	JLIM p[3]
MS	rs12749591	*PRKCZ*	-	-	-	-	2.4×10^{-4}	$\mathbf{4 \times 10^{-4}}$
MS	rs35967351	*SLAMF7*	2.3×10^{-3}	0.13	4.1×10^{-3}	0.79	5.4×10^{-8}	$\mathbf{<10^{-6}}$*
MS	rs35967351	*NHLH1*	8.2×10^{-5}	$\mathbf{8 \times 10^{-6}}$*	-	-	8.9×10^{-3}	0.92
MS	rs1359062	*RGS1*	-	-	1.6×10^{-21}	$\mathbf{<10^{-6}}$*	-	-
MS	rs9989735	*SP140*	7.5×10^{-13}	1.00	7.0×10^{-3}	0.33	1.3×10^{-9}	$\mathbf{<10^{-6}}$*
MS	rs71624119	*ANKRD55*	2.0×10^{-10}	$\mathbf{9 \times 10^{-6}}$*	-	-	-	-
MS	rs71624119	*IL6ST*	5.9×10^{-5}	$\mathbf{3 \times 10^{-4}}$	4.9×10^{-4}	0.77	-	-
MS	rs4912804	*PCDHB13*	1.4×10^{-3}	$\mathbf{8 \times 10^{-4}}$	-	-	-	-
MS	rs917116	*JAZF1*	6.2×10^{-16}	$\mathbf{<10^{-6}}$*	3.3×10^{-6}	0.89	1.1×10^{-3}	0.99
MS	rs60600003	*ELMO1*	1.2×10^{-8}	$\mathbf{1 \times 10^{-6}}$*	1.5×10^{-4}	0.006	-	-
MS	rs1966115	*ZC2HC1A*	4.5×10^{-12}	$\mathbf{1 \times 10^{-6}}$*	3.4×10^{-40}	$\mathbf{<10^{-6}}$*	3.4×10^{-30}	$\mathbf{<10^{-6}}$*
MS	rs1966115	*PKIA*	1.2×10^{-15}	1.00	-	-	1.1×10^{-9}	$\mathbf{6 \times 10^{-6}}$*
MS	rs34383631	*CD5*	2.7×10^{-3}	$\mathbf{1.3 \times 10^{-3}}$	-	-	4.5×10^{-3}	0.74
MS	rs10783847	*METTL21B*	8.8×10^{-21}	$\mathbf{<10^{-6}}$*	2.0×10^{-21}	$\mathbf{<10^{-6}}$*	-	-
MS	rs7132277	*CDK2AP1*	2.1×10^{-4}	0.57	1.3×10^{-8}	0.025	6.5×10^{-13}	$\mathbf{<10^{-6}}$*
MS	rs7132277	*ATP6V0A2*	1.1×10^{-3}	$\mathbf{1.1 \times 10^{-3}}$	-	-	-	-
MS	rs12946510	*GSDMA*	6.1×10^{-8}	$\mathbf{7 \times 10^{-4}}$	-	-	-	-
MS	rs12946510	*GSDMB*	4.1×10^{-17}	$\mathbf{<10^{-6}}$*	-	-	2.8×10^{-4}	0.85
MS	rs12946510	*ORMDL3*	5.7×10^{-13}	$\mathbf{<10^{-6}}$*	-	-	3.1×10^{-26}	$\mathbf{<10^{-6}}$*
MS	rs17785991	*SLC9A8*	2.2×10^{-6}	$\mathbf{2 \times 10^{-4}}$	6.3×10^{-9}	1.00	5.2×10^{-3}	0.87
IBD	rs13001325	*IL18R1*	7.2×10^{-11}	$\mathbf{5 \times 10^{-6}}$*	-	-	-	-
IBD	rs3749171	*GPR35*	7.6×10^{-3}	0.76	9.5×10^{-8}	$\mathbf{2 \times 10^{-4}}$	2.1×10^{-3}	0.57
IBD	rs55770741	*ERAP2*	2.2×10^{-60}	$\mathbf{<10^{-6}}$*	1.9×10^{-57}	$\mathbf{<10^{-6}}$*	1.1×10^{-105}	$\mathbf{<10^{-6}}$*
IBD	rs17622378	*SHROOM1*	-	-	1.6×10^{-3}	$\mathbf{2 \times 10^{-4}}$	-	-

Disease	Lead SNP[1]	Gene	CD4+ T cell		CD14+ monocytes		LCL	
			eQTL p[2]	JLIM p[3]	eQTL p[2]	JLIM p[3]	eQTL p[2]	JLIM p[3]
IBD	rs17622378	KIF3A	4.0×10^{-5}	$\mathbf{4\times10^{-5}}$	-	-	-	-
IBD	rs181826	PCDHB13	1.4×10^{-3}	$\mathbf{7\times10^{-4}}$	-	-	-	-
IBD	rs444210	RNASET2	9.3×10^{-51}	$\mathbf{<10^{-6*}}$	3.9×10^{-14}	1.00	9.0×10^{-8}	1.00
IBD	rs7848647	TNFSF15	1.3×10^{-3}	0.007	1.3×10^{-16}	$<10^{-6*}$	-	-
IBD	rs34779708	CUL2	2.7×10^{-7}	$\mathbf{4\times10^{-5}}$	-	-	4.1×10^{-3}	0.72
IBD	rs2590348	CISD1	-	-	9.0×10^{-13}	$<10^{-6*}$	-	-
IBD	rs12448902	TUFM	1.0×10^{-20}	$\mathbf{<10^{-6*}}$	5.6×10^{-28}	$<10^{-6*}$	-	-
IBD	rs4795397	RARA	1.5×10^{-3}	$\mathbf{3\times10^{-4}}$	-	-	-	-
IBD	rs9808651	ETS2	-	-	1.4×10^{-7}	$\mathbf{2\times10^{-4}}$	-	-
IBD	rs4456788	ICOSLG	2.9×10^{-6}	1.00	1.4×10^{-6}	$\mathbf{1.1\times10^{-4}}$	6.8×10^{-4}	0.97
IBD	rs2266961	UBE2L3	1.0×10^{-4}	$\mathbf{8\times10^{-4}}$	1.0×10^{-9}	$<10^{-6*}$	-	-
IBD	rs2143178	RPL3	4.1×10^{-4}	$\mathbf{5\times10^{-4}}$	5.8×10^{-3}	0.87	-	-
Crohn	rs6752107	scaRNA5	1.6×10^{-4}	0.11	1.8×10^{-20}	$<10^{-6*}$	-	-
Crohn	rs71624119	ANKRD55	2.0×10^{-10}	$\mathbf{2\times10^{-5*}}$	-	-	-	-
Crohn	rs71624119	IL6ST	5.9×10^{-5}	$\mathbf{4\times10^{-4}}$	4.9×10^{-4}	0.78	-	-
Crohn	rs3801810	SKAP2	9.7×10^{-14}	$\mathbf{<10^{-6*}}$	7.8×10^{-5}	1.00	-	-
UC	rs2147905	TNFRSF14	7.9×10^{-3}	0.25	-	-	2.5×10^{-8}	$\mathbf{4\times10^{-5*}}$
UC	rs11742304	TPPP	1.0×10^{-4}	1.00	-	-	1.5×10^{-7}	$\mathbf{9\times10^{-6*}}$
UC	rs4728142	TNPO3	1.2×10^{-5}	**0.003**	4.0×10^{-8}	0.85	-	-
UC	rs11150589	ITGAL	4.2×10^{-12}	$\mathbf{2\times10^{-5*}}$	6.8×10^{-4}	0.96	-	-
UC	rs889561	NFAT5	5.6×10^{-3}	0.32	-	-	7.7×10^{-3}	$\mathbf{1.9\times10^{-4*}}$
UC	rs889561	ZFP90	5.9×10^{-23}	$\mathbf{1\times10^{-6*}}$	2.3×10^{-19}	$<10^{-6*}$	5.0×10^{-3}	0.85
UC	rs6017342	OSER1	2.8×10^{-3}	**0.003**	-	-	-	-
RA	rs4681851	FLNB	1.1×10^{-6}	$\mathbf{2.2\times10^{-4}}$	3.1×10^{-3}	0.89	1.3×10^{-3}	0.78
RA	rs71624119	ANKRD55	2.0×10^{-10}	$\mathbf{9\times10^{-6*}}$	-	-	-	-
RA	rs71624119	IL6ST	5.9×10^{-5}	$\mathbf{2\times10^{-4}}$	4.9×10^{-4}	0.77	-	-
RA	rs3807306	IRF5	-	-	2.0×10^{-4}	0.97	6.7×10^{-20}	$<10^{-6*}$
RA	rs12936409	NR1D1	1.8×10^{-3}	$\mathbf{3\times10^{-4}}$	-	-	-	-

Disease	Lead SNP[1]	Gene	CD4+ T cell		CD14+ monocytes		LCL	
			eQTL p[2]	JLIM p[3]	eQTL p[2]	JLIM p[3]	eQTL p[2]	JLIM p[3]
RA	rs12936409	RARA	1.5×10^{-3}	$\mathbf{5 \times 10^{-4}}$	-	-	-	-
CEL	rs1359062	RGS1	-	-	1.6×10^{-21}	$\mathbf{< 10^{-6}}$*	-	-
CEL	rs1980422	CD28	1.2×10^{-3}	$\mathbf{1.6 \times 10^{-3}}$	-	-	-	-
CEL	rs2097282	CCR2	3.0×10^{-5}	$\mathbf{6 \times 10^{-5}}$*	1.7×10^{-8}	1.00	-	-
CEL	rs79758729	ELMO1	1.2×10^{-8}	$\mathbf{3 \times 10^{-6}}$*	4.0×10^{-4}	0.07	-	-
CEL	rs1893592	UBASH3A	4.5×10^{-14}	$\mathbf{4 \times 10^{-6}}$*	-	-	5.2×10^{-6}	1.00
CEL	rs4821124	UBE2L3	1.0×10^{-4}	$\mathbf{1.3 \times 10^{-3}}$	1.0×10^{-9}	$\mathbf{3 \times 10^{-5}}$*	-	-
T1D	rs12416116	KLLN	-	-	-	-	4.7×10^{-4}	$\mathbf{9 \times 10^{-5}}$
T1D	rs917911	CLEC2B	1.4×10^{-5}	$\mathbf{1.4 \times 10^{-5}}$*	6.8×10^{-5}	0.83	1.6×10^{-3}	0.64
T1D	rs705705	SUOX	9.3×10^{-6}	$\mathbf{6 \times 10^{-6}}$*	1.9×10^{-10}	0.98	9.3×10^{-3}	0.044
T1D	rs72727394	RASGRP1	3.9×10^{-3}	0.72	-	-	3.7×10^{-10}	$\mathbf{9 \times 10^{-4}}$
T1D	rs7239671	CD226	2.1×10^{-3}	0.83	9.7×10^{-3}	0.13	4.0×10^{-3}	$\mathbf{5 \times 10^{-4}}$
T1D	rs280497	KEAP1	-	-	6.7×10^{-3}	0.82	2.0×10^{-4}	$\mathbf{1.6 \times 10^{-4}}$
T1D	rs280497	CDC37	-	-	8.3×10^{-3}	0.85	1.4×10^{-3}	$\mathbf{1.0 \times 10^{-3}}$
T1D	rs6518350	LINC01424	-	-	-	-	9.0×10^{-4}	$\mathbf{5 \times 10^{-4}}$

[1] Variant with the minimum association p value to disease in the ImmunoChip summary statistics.

[2] Minimum eQTL p value for any SNP within 100kb of the lead SNP. Dashes (−) indicate genes that are either not detected or with minimum eQTL $p > 0.01$ in that cell type.

[3] Highlighted in bold are disease-eQTL pairs with false discovery rate $< 5\%$.

Asterisk (*) marks eQTL genes passing Bonferroni correction.