Improved chromatin QTL mapping with CACTI

Lili Wang[1,2], Xuanyao Liu[1,2,3*]

[1]The Committee on Genetics, Genomics and Systems Biology, University of Chicago, Chicago, IL, USA
[2]Department of Medicine, Section of Genetic Medicine, University of Chicago, Chicago, IL, USA
[3]Department of Human Genetics, University of Chicago, Chicago, IL, USA

*Correspondence to xuanyao@uchicago.edu

SUMMARY

Mapping chromatin quantitative trait loci (cQTLs) is crucial for elucidating the regulatory mechanisms governing gene expression and complex traits. However, current cQTL mapping methods suffer from limited detection power, particularly at existing sample sizes, and are constrained by peak-calling accuracy. To address these limitations, we present CACTI, a novel method that improves cQTL mapping by leveraging correlations between neighboring regulatory elements. Across diverse histone marks (H3K4me1, H3K4me3, H3K27ac, H3K27me3 and H3K36me3) and cell types, CACTI identifies 51%-255% more cQTL signals compared to conventional single-peak-based approaches. Using CACTI, we generate a comprehensive cQTL map for the five histone marks across multiple cell types and perform colocalization analyses with GWAS loci from 44 complex traits. CACTI cQTLs colocalize with 6%-47% of GWAS loci, which is on average 15%-424% more than the standard cQTL mapping method across different marks. 24%-75% of colocalized GWAS loci show no colocalization with eQTLs. This underscores CACTI's unique ability to uncover regulatory mechanisms that would otherwise remain undetected by eQTL analysis alone, significantly improving the functional interpretation of GWAS findings.

INTRODUCTION

The majority of disease-associated loci identified by genome-wide association studies (GWAS) are found in non-coding regions of the genome[1]. They have been found to be enriched for genetic variants influencing gene regulation (i.e., expression quantitative trait loci or eQTLs), suggesting their regulatory functions[2]. Yet, variations in gene expression only account for a small proportion of trait heritability[3–6]. The missing link has led to a growing interest in exploring additional molecular traits to better understand the regulatory mechanisms underlying GWAS loci. Of particular interest is the regulatory impact of chromatin structure (i.e., histone modifications and accessibility).

Recent advancements in high-throughput assays measuring chromatin structure have opened up new opportunities for studying molecular mechanisms of complex traits[7]. Many studies have identified variants influencing the chromatin structure of regulatory elements (chromatin quantitative trait loci or cQTLs) throughout the genome[8–12]. Additionally, studies have observed "priming effects" of cQTLs[13–15], where genetic effects on nearby regulatory elements measured in a steady state can capture context-dependent changes in gene expression. For example, one study showed that cQTLs can reveal trait-associated variants that are not observed as eQTLs, likely obscured by not using the correct contexts or states[15]. Collectively, mapping cQTLs can provide critical insights into how genetic variation impacts chromatin state, gene regulation, and complex trait.

Nevertheless, two challenges remain and compromise cQTLs detection. First, the sample sizes in most epigenetic studies are often limited[9,14], reducing the statistical power to detect cQTLs. Second, accurate peak calling is difficult for certain histone modifications[16], particularly those covering broad or mixed regions with lowly enriched read coverage[17]. These modifications often yield inconsistent peak calls depending on the performance of peak callers[16]. These challenges hinder our ability to capture a complete map of cQTLs and underscore the need for more powerful and robust methods to improve detection power and accuracy in cQTL mapping.

To address the challenges, we introduce ChromAtin quantitative loCi of mulTIple peaks (CACTI), a method that identifies cQTLs by leveraging correlations among nearby cis-regulatory elements. Studies show that regulatory elements in close proximity often exhibit coordinated regulation and share genetic control, likely influenced by the three-dimensional organization of the genome[18,19]. CACTI incorporates these correlations into the cQTL mapping using a multivariate association test[20], which detects cQTLs of coregulated elements jointly. Multivariate association methods tend to be more powerful than traditional univariate methods, enhancing cQTL detection power and uncovering broader genetic influences on chromatin regulation. Additionally, we introduce segment-based–CACTI (CACTI-S), which skips peak calling entirely and maps cQTLs directly, eliminating the bias of peak calling introduced for certain peak structures.

We applied CACTI to a collective dataset of various epigenetic marks (H3K27ac, H3K4me1, H3K4me3, H3K36me3, and H3K27me3) in several cell types and tissues (LCL, macrophage,

brain, heart, lung, and muscle) and constructed a comprehensive map of cQTLs associated. CACTI consistently outperformed traditional mapping methods of single-peak–based mapping for marks with either narrow or broad peak signatures. Furthermore, our map of cQTLs provided insights into the regulatory mechanisms underlying trait-associated loci.

RESULTS

Overview of the method

We developed a method, called chromatin quantitative loci of multiple peaks (CACTI), to increase the power of chromatin QTLs (cQTLs) detection. In contrast to the standard cQTL calling method that tests the association between single nucleotide polymorphisms (SNPs) and individual peaks, CACTI leverages the correlation structure among proximal peaks[18,19] by testing the association between multiple peaks and SNPs using multivariate association tests[20], thereby increasing the power to detect cQTLs. The method consists of two main steps (Figure 1).

First, CACTI groups neighboring peaks in a fixed-width window. The window size may vary depending on the histone marks or chromatin features being analyzed, though window sizes do not significantly affect cQTL detection power (Supplemental Notes). Next, CACTI tests associations between *cis* SNPs (i.e., within a close flanking region of window) and multiple peaks grouped in each window using a powerful and robust Principal Component (PC)-based omnibus test (PCO)[20,21]. PCO is a multivariate test that combines six PC-based statistical tests: PCMinP, PCFisher, PCLC, WI, Wald and VC (Methods). Each test combines multiple PCs differently, which allows signals under various genetic architectures to be captured. For windows containing only one peak, we used the standard univariate association test to obtain p-values. To adjust for multiple testing, we used Storey's q-value[22] to estimate false discovery rate (FDR) across all windows tested, regardless of the number of peaks contained within each window.

For histone marks that have broad peak signatures and low signal-to-noise ratios (such as H3K36me3 or H3K27me3), it can be difficult to accurately call peaks and map cQTLs[16]. We therefore developed segment-based–CACTI (CACTI-S), which maps cQTLs directly without peak calling. Specifically, instead of calling peaks, we divided the genome into small non-overlapping segments (e.g., 5kb), filtered out the low-quality segments with extremely low counts of aligned reads, and grouped the remaining segments using non-overlapping fixed-width windows (Methods). Similar to the standard CACTI, we used PCO to test the association between SNPs and window regions containing multiple segments. CACTI-S leverages correlations between neighboring segments and further increases detection power by mitigating the impact of inaccurate peak calling on cQTL mapping.

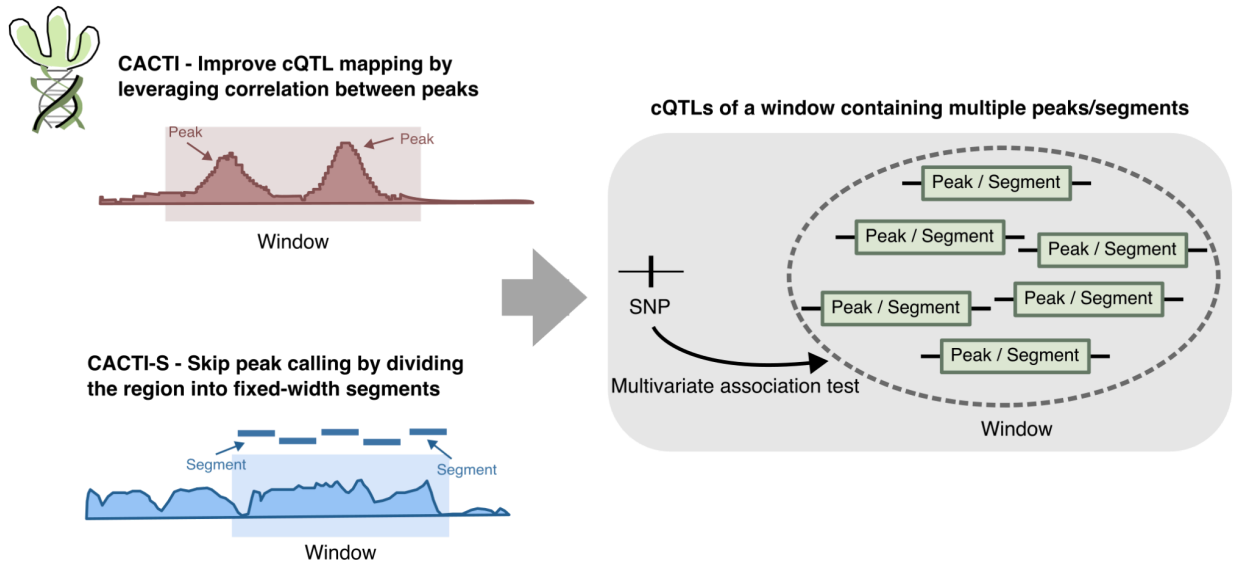

Figure 1. Overview of CACTI. CACTI groups physically adjacent peaks within a regulatory region using a fixed-width window and tests the association between each SNP and the peak group jointly using the multivariate association test. The multivariate association approach leverages the correlation among the peaks and improves mapping power. CACTI-S uses the same principles (i.e., multivariate association tests), but divides the regulatory region into fixed-width segments and then groups them into a window, eliminating any reliance on peak calling.

CACTI substantially outperforms single-peak–based mapping

To evaluate the performance of CACTI, we analyzed ChIP-seq data of the histone mark H3K27ac from lymphoblastoid cell lines (LCLs) of 78 Yoruba individuals[9,23]. Peaks were called following the procedure described in Fair et al.[23] (Methods). We found a total of 100,844 peaks, which we grouped into non-overlapping windows. We experimented with different window sizes (10kb, 25kb, and 50kb) and found CACTI is not sensitive to the choice of window sizes (Supplementary Notes, Figure S1); therefore, we used 50kb as the window size, which resulted in 1–21 peaks per window. In total, CACTI identifies 7027 windows (25% of total tested windows) that have at least one cQTLs (5% FDR), which we refer to as cwindows.

We compared CACTI to the standard single-peak–based method implemented by QTLtools[23,24] by examining the overlap between cwindows detected by CACTI and peaks with cQTLs (cpeaks) identified by single-peak mapping. CACTI identifies significantly more cQTL signals than the single-peak method: single-peak–based method detects 4967 cpeaks, 96% of which map to 3577 cwindows detected by CACTI; CACTI detects an additional 3450 cwindows (49% of total cwindows; Figure 2A). Notably, CACTI tests smaller cis regions compared to the single-peak univariate test, as it requires the cis variants to be shared by all peaks in the window (Methods). To verify the cQTLs uniquely detected by CACTI, we examined the enrichment of these cQTLs in blood eQTLs[25], reasoning that genetic variants regulating region

accessibility are expected to influence gene expression levels. Indeed, we found the CACTI-specific cQTLs are 1.39-fold enriched in eQTLs (p-value = 1.1e-12; Figure S2), suggesting that the additional cQTLs are high quality.

CACTI outperforms the single-peak mapping primarily by capturing signals with weaker effects. Overall, the univariate association z-scores of CACTI signals are smaller than those of single-peak mapping signals at the same 5% FDR level (comparison p-value = 2.2e-16; Figure S3). For example, in a window on chromosome 5 containing four peaks, CACTI identifies cQTLs associated with this window region (lead p-value = 8.35e-11), while single-peak mapping identifies no significant cQTLs (p-values = 1.13e-9 to 1.43e-4 for each single peak; Figure 2B).

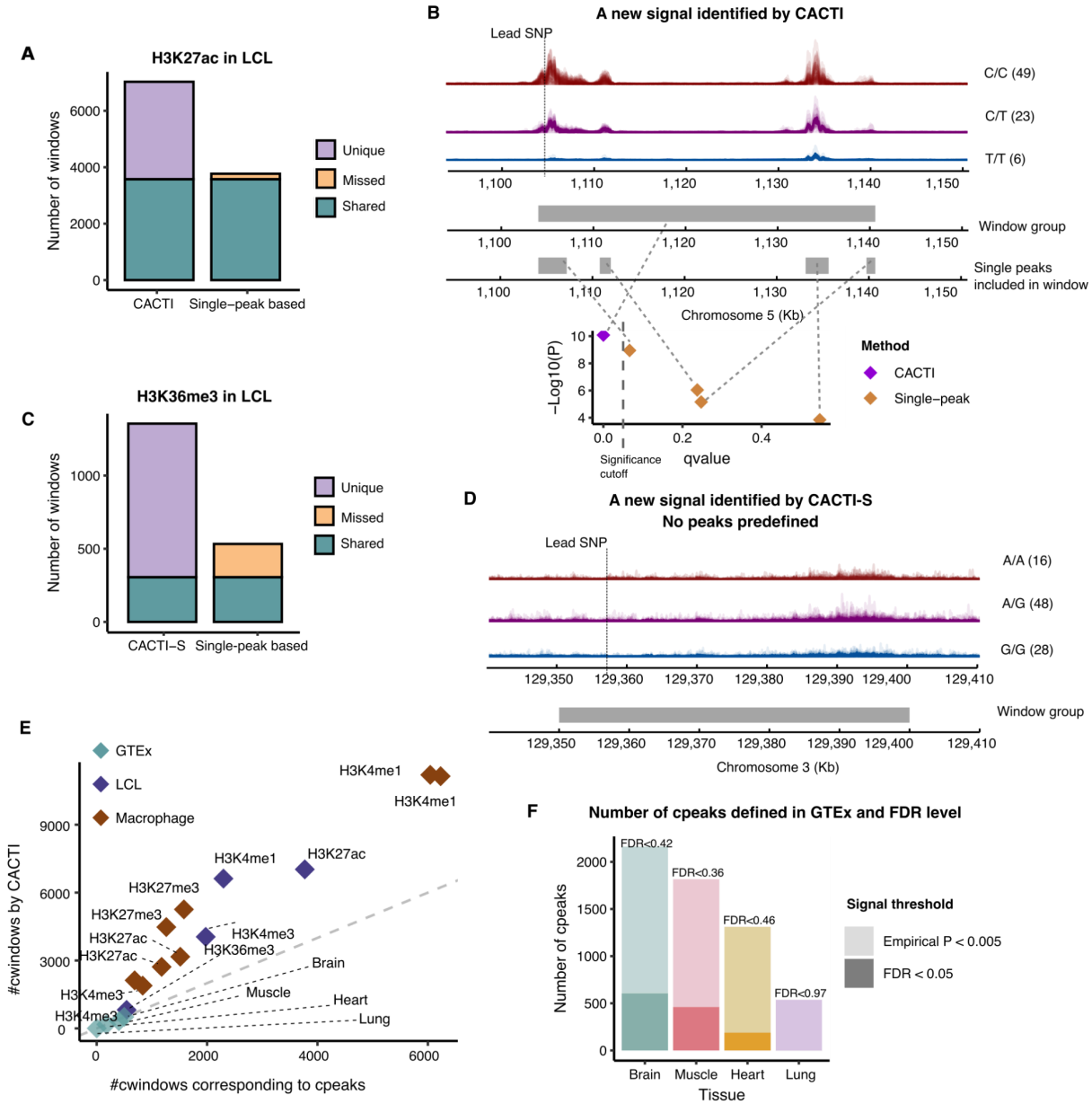

Figure 2. Performance comparison of CACTI to standard mapping methods.

(A) Comparison of signals detected by CACTI versus single-peak based mapping method for mark H3K27ac in LCL. Shared cQTLs (i.e., cwindows containing at least one cpeak) are represented in green, cQTLs specific to single-peak methods (i.e., missed by CACTI) are represented in yellow, and cQTLs specific to CACTI (i.e., missed by single-peak methods) are represented in purple.
(B) An example of a new cwindow signal discovered by CACTI that is undetected by single-peak method. The upper track displays the aligned reads for individuals classified by the genotypes of the lead cQTL in this region (marked by a dashed line). The middle track represents the cwindow region. The bottom track shows the single peaks included in the window. The scatter plot shows the p-value and q-value from association tests of the window and single peaks, respectively.
(C) Comparison of signals detected by CACTI-S versus single-peak–based mapping method for mark H3K36me3 in LCL.
(D) An example of a new cwindow signal discovered by CACTI-S.
(E) Comparison of the number of signals identified by CACTI and standard mapping methods across various datasets. Y-axis shows the number of cwindows identified by CACTI. X-axis shows the number of windows containing at least one cpeak identified by the standard mapping. Each dot represents a dataset. Color represents different cell types or tissues.
(F) The number of cpeak signals in eGTEx under the p-value cutoff used in the original paper and under a more stringent p-value cutoff. X-axis shows the four eGTEx tissues. Y-axis shows the number of cpeaks defined under two different p-value thresholds. The darker color represents cpeaks defined under FDR<0.05. We performed multiple testing correction using q-value. The lighter color represents cpeaks using the p-value threshold used in the eGTEx paper, i.e. empirical p-value < 0.005. Labels on the top of bars give the largest FDR levels corresponding to the cpeaks defined under the empirical p-value threshold.

CACTI-S skips peak calling and outperforms single-peak–based mapping for broad mark H3K36me3

Some histone marks, such as H3K36me3, have broad peak signatures of lowly enriched reads and poor signal-to-noise ratios, making the peak calling challenging and reducing the power to detect cQTLs. To address this challenge, we adapted the idea of CACTI and developed CACTI-S, which skips peak calling entirely and directly maps cQTLs by testing small segments of high-quality mapped reads grouped in fixed-sized windows (Methods).

To evaluate the performance of CACTI-S, we analyzed a CUT&TAG dataset of the histone mark H3K36me3[23]. We divided the genome into small 5kb segments, filtering out any segments with no or extremely low read alignments to reduce noise (Methods). We grouped the remaining segments into 50kb fixed-width windows. We tested the associations between the window regions and *cis* variants (within 500kb) using the multivariate approach[20] described above (Methods). In total, CACTI-S detects 1355 cwindows (4.82% of total tested windows) with at least one cQTLs at 5% FDR (Figure 2C). We compared the performance of CACTI-S with the standard single-peak based cQTL mapping, as described in Fair et al.[23] (Methods). The single-peak univariate mapping method identifies only 338 cpeaks (2.41% of the total peaks tested) associated with at least one cQTL (Figure 2C). Most of the cpeaks (246/338; 73%)

overlap with cwindows detected by CACTI-S (Figure 2C). Of the 1355 cwindows detected by CACTI-S, 1049 are CACTI-S specific signals. These cQTLs are significantly enriched in eQTLs from whole blood (OR 2.07, p-value = 1.9e-21; Figure S2).

CACTI-S attains substantially higher power than the standard univariate peak-based method, not only because multivariate associations are generally more powerful, but also because they overcome the difficulty of accurate peak calling for broad-peak signatures. For example, CACTI-S successfully captures significant cQTL signals (lead p-value = 2.68e-18) for a window on chromosome 3 with no identifiable peaks (Figure 2D). Standard single-peak based tests excluded this window from analyses. In instances where peaks are defined and tested by standard univariate association method, CACTI-S still attains higher statistical power using the multivariate approach (Figure S4); however, we note that when peak calling is accurate (e.g., H3K27ac), CACTI outperforms CACTI-S (Supplementary Notes; Figure S6). In summary, CACTI-S significantly outperforms univariate cQTL mapping for histone marks with broad peak signatures, making it a powerful tool for studying the genetic controls of these marks.

A comprehensive map of cQTLs across marks and cell types

We applied CACTI and CACTI-S to datasets spanning various histone marks, cell types, and tissues to obtain a comprehensive map of cQTLs. The datasets include ChIP-seq data of H3K27ac, H3K4me1, H3K4eme3 in LCL[9] (N=78), Cut & Tag data of H3K36me3 in LCL[23] (N=95), and ChIP-seq data of H3K27ac, H3K27me3, H3K4me1, and H3K4me3 in uninfected and infected macrophages[14] (N=35). We used CACTI to analyze marks with narrow peak signatures (i.e., H3K27ac, H3K4me1 and H3K4me3) and CACTI-S for marks with broad peak signatures (i.e., H3K36me3 and H3K27me3). Consistent with previous findings[14], we found that cQTLs for marks H3K4me3, H3K4me1 and H3K27ac are highly shared (Figure S5). Across all datasets, CACTI and CACTI-S consistently outperform the single-peak–based method, identifying 51%–255% more cQTLs (Figure 2E, Figure S7). We observed significant eQTL enrichment of the CACTI- or CACTI-S–specific cQTLs (OR range = 1.03–2.07, p-value range = 4.11e-58–2.48e-01; Figure S2). CACTI and CACTI-S also replicate the majority of cQTLs (98%–100%) detected by the standard mapping methods.

We also applied CACTI to ChIP-seq data of H3K27ac in brain, heart, lung, and muscle from eGTEx[10] (N=66–113). The number of H3K27ac cQTLs in eGTEx is significantly lower than in other datasets. For example, we detected only 507 cwindows in the eGTEx brain (N = 113) versus 7027 cwindows in LCL (N = 78) and 3166 cwindows in macrophage (N = 35, Figure 2F). We note that while the number of cQTLs we report is much lower than the original eGTEx study[10], they used a loose p-value threshold of 0.005 and did not account for multiple testing (Figure 2, Figure S8). To reduce false discoveries and ensure a more accurate and fair comparison, we applied multiple testing corrections at 5% FDR to the eGTEx results. The correction reduces the cQTL signals of single peaks by an average of 83% (Figure 2F, Figure S8). In contrast to the significantly larger number of cQTLs detected in other histone mark datasets, CACTI detects comparable numbers of signals than the single-peak-based method in

eGTEx tissues under the same FDR threshold (Figure 2F). We therefore focused on LCL and macrophage cQTLs in follow-up analyses.

CACTI cQTLs explain more GWAS loci than standard cQTLs

To explore the regulatory effects of genetic variants associated with complex traits, we performed colocalization analysis for GWAS loci of 36 blood traits[26] and 8 immune diseases[27–33] with the comprehensive map of cQTLs across four marks in infected and uninfected macrophages and four marks in LCL (Figure 3 A,B, Methods).

We grouped GWAS loci into 1Mb regions, resulting in on average 361 GWAS loci constructed for each trait to perform colocalization analysis (Methods, Figure S9). We found that cQTLs of any mark or cell type colocalizes with 9 to 96 GWAS loci, corresponding to 6% to 47% of total GWAS loci (mean = 18%; Figure 3A). The proportion of GWAS loci explained by cQTLs is comparable to those explained by eQTLs[14,23], despite the much smaller sample size of histone mark datasets compared to gene expression datasets. Among cQTLs across marks and cell types, H3K4me1 in macrophage colocalizes with the largest proportion of GWAS loci (2–40 loci across traits, 1.36%–17.02% of total loci), followed by H3K27ac in LCL (2–30 loci across traits, 1.27%–10% of total loci) and H3K4me3 in LCL (1–24 loci across traits, 1.92%–10% of total loci). The cQTLs of repressive mark H3K27me3 colocalize with 1 to 24 (0.77%–7.5%) loci across traits (Figure 3B, Figure S9). H3K36me3 is associated with transcriptionally active genes, and its cQTLs colocalize with 1 to 13 (0.22%–7.02%) GWAS loci (Figure 3B, Figure S9). Over half of the GWAS loci colocalize with only one mark (Figure S10), suggesting that considering multiple marks offers complementary insights on explaining the mechanism of GWAS loci.

We next asked whether GWAS loci were explained by cQTL signals identified by CACTI versus the standard single-peak–based mapping methods. Comparing macrophage cQTLs of H3K4me1 and GWAS loci across all 44 traits, single-peak-based cQTLs could explain 13 GWAS loci on average across traits (Figure 3C, Methods), among which CACTI could recover the majority of them (on average 88% across traits, Figure 3C). CACTI cQTLs could additionally colocalize with 25%-250% (79% on average across traits) more than single-peak-based cQTLs in explaining GWAS loci (Figure 3C). CACTI cQTLs of H3K27ac colocalize with on average 94% more GWAS loci than single-peak–based cQTLs, and CACTI cQTLs of H3K4me3 colocalized with on average 182% more. Comparing macrophage cQTLs of H3K27me3 and GWAS loci, nearly all GWAS loci colocalization (69%-100%) are only explained by CACTI-S, highlighting the power of CACTI-S over the standard univariate method and the significance of using CACTI-S in studying the role of H3K27me3 in traits and diseases. Similar comparison for cQTLs in LCL can be found in Figure S9.

Overall, CACTI significantly enhances cQTL detection and GWAS colocalization compared to the standard single-peak-based method. This improvement facilitates the functional interpretation of GWAS loci and helps prioritize putative regulatory regions. For instance, a locus associated with monocyte percentage colocalizes with H3K27ac cQTLs in macrophages (PP4 = 0.72; Figure 3D). While the region contains two distinct peaks, the univariate cQTLs of neither

peak exhibits significant colocalization (PP4 = 0.45 and 0.077; Figure S11). Additionally, we detected colocalization (PP4 > 0.5) between this GWAS locus and eQTLs of three genes (*NCOR1*, *CENPV* and *ZSWIM7*, Methods), suggesting these may be *cis*-regulatory targets of H3K27ac-marked enhancers (Figure S11). Notably, *NCOR1* (PP4=0.93) is a key regulator of macrophage-mediated inflammation and immune responses[34], further supporting its potential role in monocyte biology. As another example, we identified a multiple sclerosis (MS)-associated locus that colocalizes with H3K4me3 cQTLs (PP4=0.55; lead p-value = 0.55; Figure 3E). The corresponding cwindow contains two distinct peaks, neither of which show significant colocalization with the GWAS locus (PP4 = 0.35 and 0.078; Figure S12). The same GWAS locus also colocalizes with eQTLs of three genes, including *TBKBP1*, *NPEPPS*, and *RP11-580I16.2* (Figure S12). Among these, *TBKBP1* (PP4 = 0.77), a gene implicated in the TNF-α/NF-κB signaling pathway, has previously been linked to MS susceptibility[35–37]. These findings suggest that the GWAS locus may influence promoter activity, as marked by H3K4me3, thereby modulating *TBKBP1* expression. While functional validation is needed to confirm the precise mechanisms, our CACTI cQTL analysis highlights plausible regulatory elements and provides a functional link between GWAS loci and disease biology.

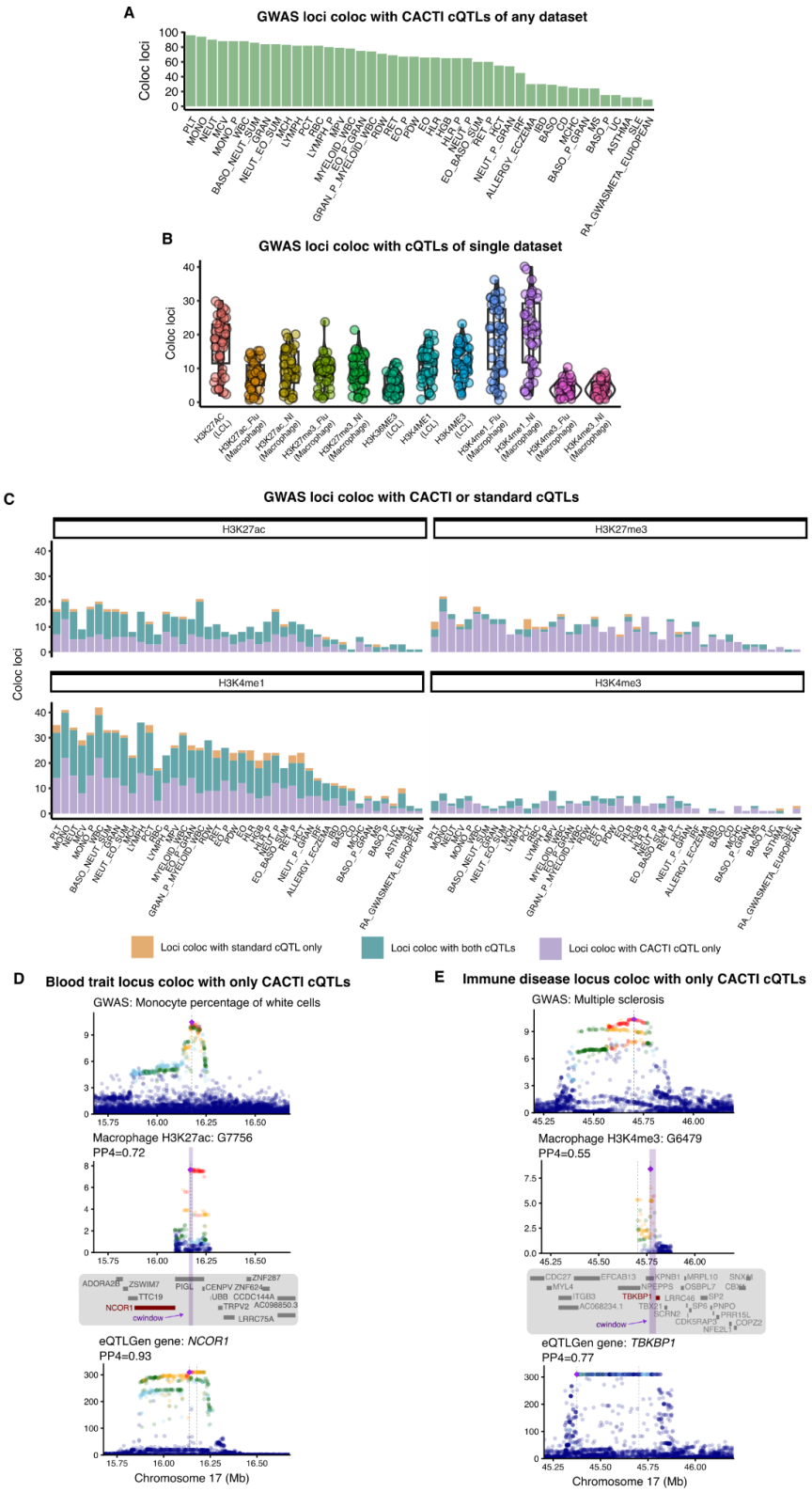

Figure 3. Colocalization of GWAS loci of 36 blood related traits and 8 immune diseases with cQTLs across marks and cell types.
(A) GWAS loci coloc with cQTLs detected by CACTI. X-axis shows 36 blood related traits and 8

immune diseases used for colocalization with cQTLs. Y-axis shows the number of GWAS loci coloc with cQTLs of any mark in any cell type. Traits are ordered by the total number of colocalized loci.
(B) GWAS loci coloc with cQTLs in single datasets detected by CACTI. X-axis shows each individual dataset of a mark and a cell type. Y-axis shows the number of GWAS loci coloc with cQTLs of the corresponding dataset. Each point represents a trait.
(C) GWAS loci colocalize with cQTLs detected by CACTI and standard mapping methods in macrophage. The X-axis shows 36 blood-related traits and eight immune diseases used for colocalization with cQTLs. The Y-axis shows the number of coloc loci. Purple bars show the loci colocalized with only CACTI cQTLs. Yellow bars show the loci that colocalize with only standard cQTLs. Blue bars show the loci colocalized with both types of cQTLs. Traits are in the same order as (A).
(D) Colocalization between a GWAS locus of monocyte percentage and CACTI cQTLs of H3K27ac in macrophage. Region is marked by purple shade. The colocalization of the GWAS locus with eQTLs of *NCOR1* is also shown.
(E) Colocalization between a GWAS locus of multiple sclerosis and CACTI cQTLs of H3K4me3 in macrophage. Region is marked by purple shade. The colocalization of the GWAS locus with eQTLs of *TBKBP1* is also shown.

CACTI cQTLs provide functional interpretation of GWAS loci beyond eQTLs

We next investigated how many GWAS loci colocalize with CACTI cQTLs but not eQTLs. Ideally, the eQTLs and cQTLs should be from the same cell type for a fair comparison. However, gene expression datasets of macrophages or LCLs typically have small sample sizes, which could limit the power of colocalization. We found that eQTLs from substantially larger gene expression datasets, such as eQTLGen (N>30,000), can better recover colocalization signals, even if cell types are not perfectly matched (Figure 4A, Supplementary Notes). To demonstrate this, we used cQTLs of LCL, eQTLs of GTEx LCL[38] (N=117) and eQTLs of eQTLGen (whole blood)[25]. For all GWAS loci that colocalize with LCL cQTLs, very few colocalize with only LCL eQTLs (Figure 4A). In contrast, eQTLs from eQTLGen recover most loci that colocalize with LCL eQTLs. We therefore used eQTLGen eQTLs to colocalize with trait and disease loci.

Overall, among all GWAS loci that colocalize with cQTLs of any mark in any cell type, 24% to 75% (4 to 38 loci) did not colocalize with eQTLs of any genes in eQTLGen (Figure 4B, Figure S13), despite the much larger sample sizes of the eQTL dataset. In one example, a GWAS locus of monocyte count on chromosome 4 colocalizes with cQTLs of H3K27ac in macrophage (PP4=0.92, Figure 4C). The GWAS locus does not colocalize with any eQTLs. A gene in this region, *IRF2*, is found to be essential for the development of non-classical monocytes, by enabling their differentiation from classical monocytes[39]. However, PP4 is only 0.0027 for the eQTLs of this gene and the GWAS locus. In another example, a GWAS locus of multiple sclerosis on chromosome 6 colocalizes with cQTLs of H3K4me1 (PP4 = 0.91, Figure 4D), but not with any eQTLs despite being near *TNFAIP3*, who is considered a central gatekeeper in inflammation and whose reduced expression can lead to increased inflammation contributing to MS[40]. It is possible that GWAS variants influence gene expression under specific context or in specific cell types, and the colocalization could not be captured in steady-state gene expression

or the wrong cell type. In contrast, cQTLs can capture the colocalization even in slightly different cell states or cell types, a result concordant with recent findings[15]. In summary, CACTI and CACTI-S cQTLs improve the interpretability of GWAS loci, by increasing the number of colocalized loci and reducing the sensitivity to cellular context, highlighting the importance of considering molecular traits beyond gene expression.

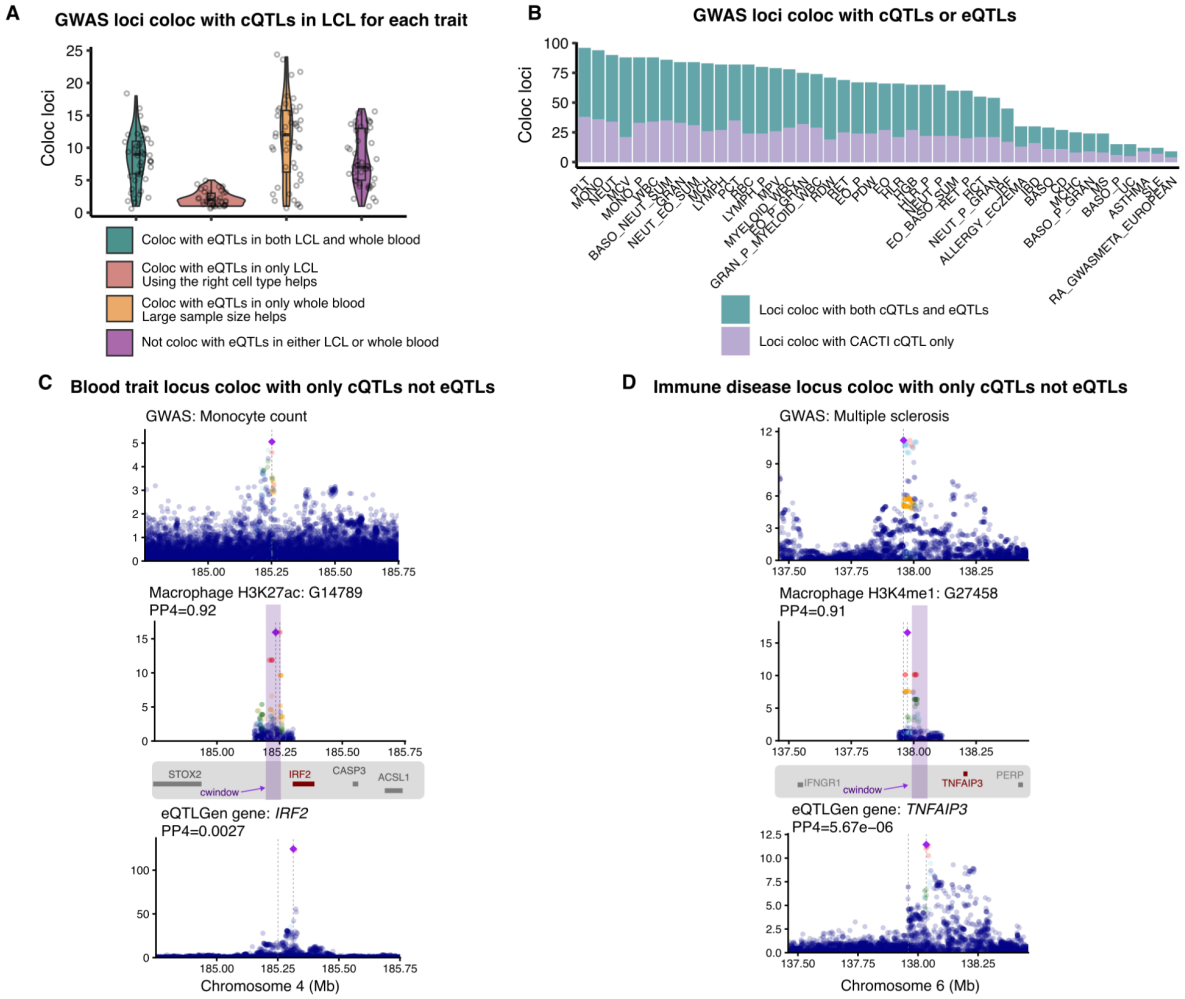

Figure 4. Colocalization of GWAS loci of 36 blood related traits and 8 immune diseases with cQTLs versus eQTLs.

(A) GWAS loci colocalize with cQTLs in LCL and eQTLs from LCL or whole blood. Each point is a blood trait or immune disease. Y-axis shows the number of loci coloc with cQTLs. These loci are categorized into four groups, including loci also coloc with eQTLs in (1) both whole blood and LCL, (2) only whole blood, (3) only LCL, (4) neither whole blood or LCL. eQTLs in whole blood are from eQTLGen. eQTLs in LCL are from GTEx.

(B) GWAS loci colocalize with only CACTI cQTLs but not eQTLs (purple) or with both cQTLs and eQTLs (blue). eQTLs are from eQTLGen.

(C) A GWAS locus monocyte count on chromosome 4 colocalizes with cQTLs of H3K27ac in macrophage identified by CACTI. This locus does not colocalize with eQTLs of any nearby genes, including *IRF2*.

(D) A GWAS locus of multiple sclerosis on chromosome 6 colocalizes with cQTLs of H3K4me1 in macrophage identified by CACTI. This locus does not colocalize with eQTLs of any nearby genes, including *TNFAIP3*.

DISCUSSION

We introduced CACTI, a powerful method for identifying cQTLs of multiple nearby peaks using multivariate association tests. CACTI and CACTI-S consistently outperform the standard single–peak–based method by 51%–255% across all histone marks in different cell types (Figure 2E). Using CACTI, we constructed a comprehensive map of cQTLs associated with five epigenetic marks across cell types. CACTI cQTLs can explain between 9 and 96 GWAS loci of 44 complex traits, corresponding to 6% to 47% of total GWAS loci (Figure 3A & B). This is comparable to the proportion of GWAS loci explained by eQTLs, even though cQTLs are mapped at a significantly smaller sample size than eQTLs. Furthermore, among all GWAS loci that colocalize with cQTLs, 24% to 75% did not colocalize with eQTLs of any genes (Figure 4B), despite the much larger sample sizes of eQTLs. The priming effects of epigenetic marks can explain some of these signals: the genetic effects on gene expression cannot be captured by steady-state gene expression or expression in the wrong cell type, but the genetic effects on epigenetic marks can be observed. These findings support the importance of assaying histone modifications and chromatin accessibility, in addition to gene expression data, as well as the value of CACTI in improving the interpretation of GWAS loci.

CACTI-S skips peak calling and significantly simplifies cQTL mapping for histone marks with broad peak signatures and low signal-to-noise ratios. The advantage of CACTI-S is especially evident when explaining GWAS loci with cQTLs. For example, almost all H3K27me3 cQTLs that colocalize with GWAS loci are mapped by CACTI-S (Figure 3C). Nonetheless, we note that CACTI is still superior to CACTI-S when mapping cQTLs in histone marks with narrow peak signatures. For example, while CACTI-S outperforms the single-peak-based method for H3K27ac in LCL, CACTI outperforms both methods (Supplementary Notes). Therefore, we recommend peak calling followed by CACTI in histone marks with a narrow peak signature.

The complete list of cQTLs generated by CACTI provides a valuable resource for further exploration of epigenetic marks and regulatory elements in molecular and complex traits. Further analyses linking the colocalized peaks to nearby genes are beyond the scope of this work, but can be accomplished through experimental approaches or by using existing resources such as ABC scores[41].

Limitations of the study

Our approach has several limitations. First, CACTI provides an effective method to group multiple peaks and map their cQTLs. Our analyses demonstrated that the performance of CACTI was not sensitive to the choices of different window sizes; however, the optimal window size could vary for different epigenetic marks. To achieve maximum power, end users can experiment with different window sizes and determine the optimal window size for their analysis. Furthermore, while we used a fixed-window in these analyses, the window size can be defined as dynamic cis-regulatory domains of various sizes[18]. Second, significant cQTLs for a cwindow containing multiple peaks do not suggest that the cQTL leads to changes in all the peaks. The associated SNP may be associated with one or more peaks in the cwindow. Combining the

analysis with univariate association effects can help identify the associated peaks as needed. Lastly, the colocalization signals of cQTL and GWAS loci do not necessarily imply that disease risk is mediated through epigenetic changes. Further studies will be needed to confirm the causal link between epigenetic changes and disease risk.

STAR METHODS

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Xuanyao Liu (xuanyao@uchicago.edu).

Materials availability

No materials were generated in the study.

Data and code availability

METHOD DETAILS

Multivariate association test

We test if a genetic variant is associated with multiple peaks grouped in a same window using the multivariate model as follows,

$$[y_1 \cdots y_K] = G[\beta_1 \cdots \beta_K] + covariates + e,$$

where G is the dosage of a reference allele representing the genotype of a SNP, β_k is the effect of the SNP on k-th peak in the window containing K peaks, and y_k is the intensity level of the k-th peak. To test if a SNP of interest is significantly associated with the window, i.e. any peaks contained in the window, we test the null hypothesis,

$$H_0\colon \beta_1 = \cdots = \beta_K = 0.$$

We use a PC-based omnibus test (PCO)[20], which is a powerful and robust PC-based approach aiming at testing the association with multiple phenotypes with no prior knowledge of the true effects.

Specifically, PCO combines multiple single PC-based tests in linear and non-linear ways, corresponding to a range of causal relationships between the genetic variant and peaks, to achieve higher power and better robustness. A single PC-based test (most commonly the first primary PC_1) is,

$$T_{PC_k} = \mu_k^T Z \sim N(\mu_k^T \beta, \lambda_k),\ 1 \leq k \leq K,$$

where Z is a $K \times 1$ vector of univariate summary statistic z-scores of the SNP for K peaks in a window, μ_k is the k-th eigenvector of the covariance matrix $\Sigma_{K \times K}$ of Z, λ_k is the corresponding eigenvalue, and β represents the true causal effect. PCO combines six PC-based tests, including,

$$PCMinP = min_{1 \le k \le K} p_k, \text{ and } PCFisher = -2 \sum_{k=1}^{K} log(p_k),$$

where p_k is the p-value of T_{PC_k}. These two tests take the best p-value of single PC-based tests and combine multiple PC p-values as the test statistic. Other tests include,

$$PCLC = \sum_{k=1}^{K} \frac{T_{PC_k}}{\lambda_k}, WI = \sum_{k=1}^{K} T_{PC_k}{}^2, Wald = \sum_{k=1}^{K} \frac{T_{PC_k}{}^2}{\lambda_k}, VC = \sum_{k=1}^{K} \frac{T_{PC_k}{}^2}{\lambda_k^2},$$

which are linear and quadratic combinations of each single PC-based test weighted by eigenvalues. The six tests achieve best power in specific genetic settings with different true causal effects. PCO takes the best p-value of the PC-based tests as the final test statistic,

$$T_{PCO} = min\, p_{\{PCMinP,\, PCFisher,\, PCLC,\, WI,\, Wald,\, VC\}},$$

to achieve robustness under unknown genetic architectures while maintaining a high power. The p-value of PCO test statistics can be computed by performing an inverse-normal transformation of the test statistics,

$$p_{T_{PCO}} = 1 - P\{min\, \Phi^{-1}(p_{\{PCMinP,\, PCFisher,\, PCLC,\, WI,\, Wald,\, VC\}}) > \Phi^{-1}(T_{PCO}^{obs})\},$$

where Φ^{-1} denotes the inverse standard normal cumulative distribution function. The p-value can be efficiently computed using a multivariate normal distribution as described in Liu et al..

To implement the PCO test, we used the summary statistics of univariate test from the tested SNP and each peak contained in the window. For the univariate test of single peaks, see section below. For more details about the input required by the multivariate association test, see Wang et al.[21].

Segment-based–CACTI (CACTI-S)

For broad marks H3K36me3 in LCL and H3K27me3 in macrophage, we applied the segment-based CACTI (CACTI-S). We first divided the whole genome into small non-overlapping segments of 5kb. We quantified the coverage of H3K36me3 and H3K27me3 reads overlapping each segment region using featureCounts[42].

We then filtered out segments with no reads or extremely low read counts to avoid including noisy regions into our analysis. To filter segments for mark H3K36me3 in LCL, we followed a similar procedure used in Fair et al.[23]. Briefly, we first normalized the raw counts per segment to counts per million (CPM) and selected the top segments with the highest abundance calculated by log2 CPM averaged across samples. To determine the number of top segments retained for downstream analysis, we calculated the number of segments that overlapped with H3K36me3 peaks. We reason that H3K36me3 peaks generally represent the transcribed regions and segments in the transcribed regions are of high quality. To filter low-quality segments for mark H3K27me3 in macrophage, we followed the procedure used in Aracena et al.[14]. Specifically, we only retained segments with aligned reads in more than 50% samples (14 samples).

To calculate the univariate association mapping between single variants and each segment, we followed the procedure of QTL mapping described in Fair et al.[23] and Aracena et al.[14]. See

details in section below. We grouped the remaining segments using non-overlapping 50kb fixed-width windows. To test the association between a window region containing multiple segments and their cis variants, we performed the multivariate association test described above. To be noted, the multivariate test requires the univariate associations between a variant and each segment contained in the window as input. Therefore, cis variants tested for multivariate association will be limited to the variants shared by all segments in the window.

QUANTIFICATION AND STATISTICAL ANALYSIS

Single-peak–based mapping

To perform the univariate cQTL mapping, i.e. the single-peak–based mapping, we followed the QTL calling procedure described in Fair et al.[23] for marks in LCL, Aracena et al.[14] for marks in macrophage, and Hou et al.[10] for H3K27ac in eGTEx.

Briefly, for marks H3K27ac, H3K4me3 and H3K4me1 in Fair et al., peak regions were called using MACS2[43], while for H3K36me3 coverage were quantified transcription across gene bodies[23]. We followed the same normalization procedure and enforced normality by applying the rank-based inverse normal transform. We used the same principal components as covariates for QTL mapping, which was implemented by QTLTools[23,24]. We used a cis window of 500kb to be consistent with the cis region used in the paper. Notably, univariate mapping tested a wider cis region compared to multivariate test, which requires the univariate association for variants shared by all peaks contained in the window. To claim cpeaks, we adjusted p-values of all peaks with Storey's q-value method[22] to account for the false discovery rate. We used the same procedure to perform the univariate association mapping for individual segments when applying CACTI-S to H3K36me3.

For marks H3K27ac, H3K27me3, H3K4me1 and H3K4me3 in Aracena et al.[14], peaks were called using MACS2. To perform univariate mapping, R package Matrix eQTL[44] was used to quantify the association between variants and peaks. Matrix eQTL was also used to obtain the univariate summary statistics of individual segments when applying CACTI-S to H3K27me3. PCs on the age and batch corrected peak matrices and PC's on genotype data were included as covariates[14]. We used cis region of 100kb for H3K27me3, to be consistent with the cis region used in the original paper.

Colocalization of GWAS loci with cQTLs and eQTLs

We performed colocalization analysis between 36 blood traits[26] and 8 immune diseases[27–33] and multiple types of molecular QTLs, including window cQTLs by the multivariate test of marks in macrophage and LCL, peak cQTLs by the univariate test of marks in macrophage and LCL, and eQTLs in eQTLGen whole blood and GTEx LCL.

To define regions to perform colocalization, we first selected the GWAS lead hit with the most significant p-value and expanded a 500kb flanking genomic region centered at the lead SNP.

We then moved on to the next most significant GWAS hit and expanded another 500kb flanking region. We stopped searching for lead SNPs when there is no GWAS hit with p-value smaller than 1e-7. Colocalization test was only performed when the GWAS loci and the molecular phenotype (window, peak, or gene) share more than 150 SNPs. We used the 'coloc.abf' function from coloc package[45] with the default priors. We defined colocalization when the molecular phenotype is a significant signal (cwindow, cpeak, or egene) and PP4 > 0.5. We used LiftOver[46] to convert genomic coordinates for GWAS loci to align them with molecular QTLs with different genome assemblies, i.e. reads in LCL datasets were aligned to the hg38 human reference genome, while reads in other datasets were based on genome version hg19. We visualized the colocalized regions using LocusCompareR[47].

ACKNOWLEDGMENTS

We thank Y. Li, Z. Liu, B. Fair, CF. NAF. and K. Aracena for helpful discussions. We thank S. Sumner for editing the manuscript. This work was completed in part with resources provided by the University of Chicago's Research Computing Center. This research was funded by the NIGMS Maximizing Investigators' Research Award (R35GM138084).

AUTHOR CONTRIBUTIONS

L.W.and X.L. developed the method and all of the analyses. L.W. implemented the method and performed data analyses under the supervision of X.L. L.W. and X.L. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science *337*, 1190–1195. https://doi.org/10.1126/science.1222794.

2. Nicolae, D.L., Gamazon, E., Zhang, W., Duan, S., Dolan, M.E., and Cox, N.J. (2010). Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. *6*, e1000888.

3. Yao, D.W., O'Connor, L.J., Price, A.L., and Gusev, A. (2020). Quantifying genetic effects on disease mediated by assayed gene expression levels. Nat. Genet. *52*, 626–633. https://doi.org/10.1038/s41588-020-0625-2.

4. Umans, B.D., Battle, A., and Gilad, Y. (2020). Where Are the Disease-Associated eQTLs? Trends Genet. *0*. https://doi.org/10.1016/j.tig.2020.08.009.

5. Mostafavi, H., Spence, J.P., Naqvi, S., and Pritchard, J.K. (2023). Systematic differences in discovery of genetic effects on gene expression and complex traits. Nat. Genet. *55*, 1866–1875. https://doi.org/10.1038/s41588-023-01529-1.

6. Liu, X., Li, Y.I., and Pritchard, J.K. (2019). Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. Cell *177*, 1022-1034.e6. https://doi.org/10.1016/j.cell.2019.04.014.

7. Park, P.J. (2009). ChIP–seq: advantages and challenges of a maturing technology. Nat. Rev. Genet. *10*, 669–680. https://doi.org/10.1038/nrg2641.

8. Farh, K.K.-H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh, N., Whitton, H., Ryan, R.J.H., Shishkin, A.A., et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature *518*, 337–343. https://doi.org/10.1038/nature13835.

9. Grubert, F., Zaugg, J.B., Kasowski, M., Ursu, O., Spacek, D.V., Martin, A.R., Greenside, P., Srivas, R., Phanstiel, D.H., Pekowska, A., et al. (2015). Genetic Control of Chromatin States in Humans Involves Local and Distal Chromosomal Interactions. Cell *162*, 1051–1065. https://doi.org/10.1016/j.cell.2015.07.048.

10. Hou, L., Xiong, X., Park, Y., Boix, C., James, B., Sun, N., He, L., Patel, A., Zhang, Z., Molinie, B., et al. (2023). Multitissue H3K27ac profiling of GTEx samples links epigenomic variation to disease. Nat. Genet., 1–12. https://doi.org/10.1038/s41588-023-01509-5.

11. Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330. https://doi.org/10.1038/nature14248.

12. McVicker, G., van de Geijn, B., Degner, J.F., Cain, C.E., Banovich, N.E., Raj, A., Lewellen, N., Myrthil, M., Gilad, Y., and Pritchard, J.K. (2013). Identification of Genetic Variants That Affect Histone Modifications in Human Cells. Science *342*, 747–749. https://doi.org/10.1126/science.1242429.

13. Alasoo, K., Rodrigues, J., Mukhopadhyay, S., Knights, A.J., Mann, A.L., Kundu, K., Hale, C., Dougan, G., and Gaffney, D.J. (2018). Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. Nat. Genet. *50*, 424–431. https://doi.org/10.1038/s41588-018-0046-7.

14. Aracena, K.A., Lin, Y.-L., Luo, K., Pacis, A., Gona, S., Mu, Z., Yotova, V., Sindeaux, R., Pramatarova, A., Simon, M.-M., et al. (2024). Epigenetic variation impacts individual differences in the transcriptional response to influenza infection. Nat. Genet. *56*, 408–419. https://doi.org/10.1038/s41588-024-01668-z.

15. Baca, S.C., Singler, C., Zacharia, S., Seo, J.-H., Morova, T., Hach, F., Ding, Y., Schwarz, T., Huang, C.-C.F., Anderson, J., et al. (2022). Genetic determinants of chromatin reveal prostate cancer risk mediated by context-dependent gene regulation. Nat. Genet. *54*, 1364–1375. https://doi.org/10.1038/s41588-022-01168-y.

16. Jeon, H., Lee, H., Kang, B., Jang, I., and Roh, T.-Y. (2020). Comparative analysis of commonly used peak calling programs for ChIP-Seq analysis. Genomics Inform. *18*, e42. https://doi.org/10.5808/GI.2020.18.4.e42.

17. Zhou, V.W., Goren, A., and Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. Nat. Rev. Genet. *12*, 7–18. https://doi.org/10.1038/nrg2905.

18. Delaneau, O., Zazhytska, M., Borel, C., Giannuzzi, G., Rey, G., Howald, C., Kumar, S., Ongen, H., Popadin, K., Marbach, D., et al. (2019). Chromatin three-dimensional interactions mediate genetic effects on gene expression. Science *364*, eaat8266. https://doi.org/10.1126/science.aat8266.

19. Gate, R.E., Cheng, C.S., Aiden, A.P., Siba, A., Tabaka, M., Lituiev, D., Machol, I., Gordon, M.G., Subramaniam, M., Shamim, M., et al. (2018). Genetic determinants of co-accessible chromatin regions in activated T cells across humans. Nat. Genet. *50*, 1140–1150. https://doi.org/10.1038/s41588-018-0156-2.

20. Liu, Z., and Lin, X. (2019). A Geometric Perspective on the Power of Principal Component Association Tests in Multiple Phenotype Studies. J. Am. Stat. Assoc. *114*, 975–990. https://doi.org/10.1080/01621459.2018.1513363.

21. Wang, L., Babushkin, N., Liu, Z., and Liu, X. (2024). *Trans*-eQTL mapping in gene sets identifies network effects of genetic variants. Cell Genomics *4*, 100538. https://doi.org/10.1016/j.xgen.2024.100538.

22. Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. U. S. A. *100*, 9440–9445. https://doi.org/10.1073/pnas.1530509100.

23. Fair, B., Buen Abad Najar, C.F., Zhao, J., Lozano, S., Reilly, A., Mossian, G., Staley, J.P., Wang, J., and Li, Y.I. (2024). Global impact of unproductive splicing on human gene expression. Nat. Genet., 1–11. https://doi.org/10.1038/s41588-024-01872-x.

24. Delaneau, O., Ongen, H., Brown, A.A., Fort, A., Panousis, N.I., and Dermitzakis, E.T. (2017). A complete tool set for molecular QTL discovery and analysis. Nat. Commun. *8*, 15452. https://doi.org/10.1038/ncomms15452.

25. Võsa, U., Claringbould, A., Westra, H.J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Yazar, S., et al. (2021). Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat. Genet. 2021 539 *53*, 1300–1310. https://doi.org/10.1038/s41588-021-00913-z.

26. Astle, W.J., Elding, H., Jiang, T., Allen, D., Ruklisa, D., Mann, A.L., Mead, D., Bouman, H., Riveros-Mckay, F., Kostadima, M.A., et al. (2016). The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. Cell *167*, 1415-1429.e19. https://doi.org/10.1016/j.cell.2016.10.042.

27. Ferreira, M.A.R., Mathur, R., Vonk, J.M., Szwajda, A., Brumpton, B., Granell, R., Brew, B.K., Ullemar, V., Lu, Y., Jiang, Y., et al. (2019). Genetic Architectures of Childhood- and Adult-Onset Asthma Are Partly Distinct. Am. J. Hum. Genet. *104*, 665–684. https://doi.org/10.1016/j.ajhg.2019.02.022.

28. Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., et al. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat. Genet. *47*, 979–986. https://doi.org/10.1038/ng.3359.

29. De Lange, K.M., Moutsianas, L., Lee, J.C., Lamb, C.A., Luo, Y., Kennedy, N.A., Jostins, L., Rice, D.L., Gutierrez-Achury, J., Ji, S.G., et al. (2017). Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nat. Genet. *49*, 256–261. https://doi.org/10.1038/ng.3760.

30. Ferreira, M.A., Vonk, J.M., Baurecht, H., Marenholz, I., Tian, C., Hoffman, J.D., Helmer, Q., Tillander, A., Ullemar, V., van Dongen, J., et al. (2017). Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. Nat. Genet. *49*, 1752–1757. https://doi.org/10.1038/ng.3985.

31. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A.P., and Price, A.L. (2018). Mixed-model association for biobank-scale datasets. Nat. Genet. *50*, 906–908. https://doi.org/10.1038/s41588-018-0144-6.

32. INTERNATIONAL MULTIPLE SCLEROSIS GENETICS CONSORTIUM (2019). Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. Science *365*, eaav7188. https://doi.org/10.1126/science.aav7188.

33. Bentham, J., Morris, D.L., Graham, D.S.C., Pinder, C.L., Tombleson, P., Behrens, T.W., Martín, J., Fairfax, B.P., Knight, J.C., Chen, L., et al. (2015). Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. Nat. Genet. *47*, 1457–1464. https://doi.org/10.1038/ng.3434.

34. Geiger, M.A., Guillaumon, A.T., Paneni, F., Matter, C.M., and Stein, S. (2020). Role of the Nuclear Receptor Corepressor 1 (NCOR1) in Atherosclerosis and Associated Immunometabolic Diseases. Front. Immunol. *11*, 569358. https://doi.org/10.3389/fimmu.2020.569358.

35. Sorosina, M., Barizzone, N., Clarelli, F., Anand, S., Lupoli, S., Salvi, E., Mangano, E., Bordoni, R., Roostaei, T., Mascia, E., et al. (2022). A multi-step genomic approach prioritized TBKBP1 gene as relevant for multiple sclerosis susceptibility. J. Neurol. *269*, 4510–4522.

https://doi.org/10.1007/s00415-022-11109-8.

36. James, T., Lindén, M., Morikawa, H., Fernandes, S.J., Ruhrmann, S., Huss, M., Brandi, M., Piehl, F., Jagodic, M., Tegnér, J., et al. (2018). Impact of genetic risk loci for multiple sclerosis on expression of proximal genes in patients. Hum. Mol. Genet. *27*, 912–928. https://doi.org/10.1093/hmg/ddy001.

37. Bouwmeester, T., Bauch, A., Ruffner, H., Angrand, P.-O., Bergamini, G., Croughton, K., Cruciat, C., Eberhard, D., Gagneur, J., Ghidelli, S., et al. (2004). A physical and functional map of the human TNF-α/NF-κB signal transduction pathway. Nat. Cell Biol. *6*, 97–105. https://doi.org/10.1038/ncb1086.

38. THE GTEX CONSORTIUM (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330. https://doi.org/10.1126/science.aaz1776.

39. O'Connor, K.W., Liu, T., Kim, S., Briseño, C.G., Georgopoulos, K., Murphy, T.L., and Murphy, K.M. (2023). Bcl6, Irf2, and Notch2 promote nonclassical monocyte development. Proc. Natl. Acad. Sci. *120*, e2220853120. https://doi.org/10.1073/pnas.2220853120.

40. Perga, S., Montarolo, F., Martire, S., Bonaldo, B., Bono, G., Bertolo, J., Magliozzi, R., and Bertolotto, A. (2020). Overexpression of the ubiquitin‐editing enzyme A20 in the brain lesions of Multiple Sclerosis patients: moving from systemic to central nervous system inflammation. Brain Pathol. *31*, 283–296. https://doi.org/10.1111/bpa.12906.

41. Fulco, C.P., Nasser, J., Jones, T.R., Munson, G., Bergman, D.T., Subramanian, V., Grossman, S.R., Anyoha, R., Doughty, B.R., Patwardhan, T.A., et al. (2019). Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. Nat. Genet. *51*, 1664–1669. https://doi.org/10.1038/s41588-019-0538-0.

42. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

43. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.

44. Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics *28*, 1353–1358. https://doi.org/10.1093/bioinformatics/bts163.

45. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLOS Genet. *10*, e1004383. https://doi.org/10.1371/journal.pgen.1004383.

46. Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F., et al. (2006). The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. *34*, D590–D598. https://doi.org/10.1093/nar/gkj144.

47. Liu, B., Gloudemans, M.J., Rao, A.S., Ingelsson, E., and Montgomery, S.B. (2019).

Abundant associations with gene expression complicate GWAS follow-up. Nat. Genet. *51*, 768–769. https://doi.org/10.1038/s41588-019-0404-0.