Supplementary Note

Contents

1 ATAC-seq in lymphoblastoid cell lines (LCLs)

1.1 Samples

We obtained 76 lymphoblastoid cell lines from Coriell (https://catalog.coriell.org/). The lines were prepared from blood samples collected in Great Britain and also genotyped in the 1000 Genomes Project. The collected samples are as follows: HG00096, HG00098, HG00100, HG00101, HG00102, HG00103, HG00105, HG00106, HG00107, HG00108, HG00109, HG00111, HG00112, HG00113, HG00114, HG00115, HG00116, HG00117, HG00118, HG00119, HG00120, HG00121, HG00122, HG00125, HG00126, HG00127, HG00128, HG00129, HG00130, HG00131, HG00132, HG00135, HG00136, HG00137, HG00138, HG00140, HG00141, HG00142, HG00143, HG00145, HG00146, HG00148, HG00149, HG00150, HG00151, HG00152, HG00154, HG00155, HG00156, HG00157, HG00158, HG00159, HG00231, HG00232, HG00233, HG00234, HG00237, HG00238, HG00239, HG00242, HG00243, HG00244, HG00245, HG00246, HG00247, HG00254, HG00255, HG00262, HG00263, HG00264, HG00265, HG01334, HG01789, HG01790, HG01791, HG02215. The remaining 24 LCL lines were previously published in [1] for a total of 100 LCL samples.

1.2 ATAC-seq protocol

ATAC-seq library preparation was performed as previously described [1].

1.3 Illumina sequencing

76 ATAC-seq libraries each prepared with one of 76 Nextera i5 and i7 tag combinations (see above), were pooled in equal volumes. Index tag ratios were assessed by a single MiSeq run. Index tag ratios were balanced according to the MiSeq run before running 44 HiSeq 2500 lanes. In combination with the ATAC-seq data for the 24 samples previously sequenced [1], we obtained a total of 4.4 billion mapped fragments (8.7 billion reads) on autosomes. All sequencing results are available from European Nucleotide Archive (ID: PRJEB9977). The ATAC-seq for GM12878 sample was additionally performed and sequenced separately.

2 Data preprocessing

2.1 Read alignment

Prior to the read alignment we performed sequencing adapter trimming by using skewer [2]. Reads were mapped to assembly GRCh37 using BWA 0.7.4 [3].Following read mapping, we selected fragments (read-pairs) that were uniquely mapped, where at least one of mate-pairs had a quality score of $>$10, aligned with 1 gap, with three base mismatches or less. Any read pairs with an insert size less than 38bp or greater than 10Kb, or on different chromosomes, were excluded from subsequent analyses.

We also used two external sequencing data sets, RNA-seq data from the gEUVADIS project [4] and CTCF binding ChIP-seq data from Ding et al. [5]. For the gEUVADIS RNA-seq data, we mapped reads to GRCh37 using Bowtie2 [6] and constructed spliced alignments using Tophat2 [7] with default settings, using gene annotation information given by Ensembl 69 as a guide for the alignment. Following read mapping, we selected fragments (read-pairs) that were uniquely mapped, where at least one of mate-pairs had a quality score of $>$10, aligned with 1 gap, with three base mismatches or less. Any read pairs with an insert size less than 75bp or greater than 500Kb, or on different chromosomes, were excluded from subsequent analyses. For the CTCF ChIP-seq data, the same alignment process to our ATAC-seq data was applied. In addition any read pairs with an insert size less than 50bp were excluded from subsequent analyses.

2.2 Peak calling for ATAC-seq and CTCF ChIP-seq data

We first pooled all samples from ATAC-seq data ($N = 100$) and CTCF ChIP-seq data ($N = 50$), respectively. For ATAC-seq data, we counted both ends of sequenced fragments (+4bp from downstream end and -4bp from upstream end) as transposase cut sites at each genome coordinate. For CTCF ChIP-seq data, we counted the midpoint of sequenced fragments (mate pairs) at each genome coordinate.

Then for each of those coverage depth data, we fitted two Gaussian kernel density estimations, one for smoothing peaks with bandwidth equal to 100bp (referred to as the peak coverage) and the other for creating background with bandwidth equal to 1kb (referred to as the background coverage). A peak was defined by comparing the two smoothed coverage depth data. We called a peak when the peak coverage was greater than the background coverage and the peak coverage in fragment per million (FPM) was greater than 0.001, that is

$$\frac{\text{(peak coverage depth)}}{\text{(total fragments sequenced in megabase)}} > 0.001.$$

We defined 277,128 peaks for our ATAC-seq data and 141,147 peaks for CTCF ChIP-seq data.

2.3 Counting fragments and FPKM (RPKM) calculation

For ATAC-seq and CTCF ChIP-seq data, we counted the number of sequenced fragments of which one or other sequenced end overlaps with the annotated peak. Likewise, for RNA-seq data, we counted the number of sequenced fragments (mate-pairs) of which one or other sequenced end overlaps with an union of annotated Ensembl gene exons.

Let Y_{ij} be the fragment count of the feature, either gene or peak, j ($j = 1, \ldots, J$) for an individual i ($i = 1, \ldots, N$). We calculated $\log_2$ FPKM (fragments per kilobase of exon per million fragments mapped), y_{ij},

for sample i at feature j as follows:

$$y_{ij} = \log_2 \left(\frac{Y_{ij} + 1}{l_j Y_i} \right),$$

where l_j is the feature length (peak length for ATAC-seq and CTCF ChIP-seq / length of an union of annotated gene exons for RNA-seq) in kilobase and $Y_i = \sum_{j=1}^{J} Y_{ij}/10^6$ is the total fragment (read) count in megabase for the individual i.

2.4 GC correction for fragment counts and FPKMs

We corrected for varying amplification efficiency of different GC contents using the method described in [8]. We first calculated GC content of each union of annotated gene exons for RNA-seq data and that of each peak for other sequencing data, which is mean G/C base counts within a feature over the feature length. Then we assigned all features to 200 approximately equally sized bins $\{\mathcal{B}_1, \ldots, \mathcal{B}_{200}\}$ based on the GC content. Let $S_{il} = \sum_{j \in \mathcal{B}_l} Y_{ij}$ be the number of fragments in bin l from individual i. For each bin, for each individual, we calculated the $\log_2$ relative enrichment, F_{il}, of fragments in each GC bin, such that

$$F_{il} = \log_2 \left(\frac{S_{il}/S_{\cdot l}}{S_{i\cdot}/S_{\cdot\cdot}} \right),$$

where $S_{\cdot l} = \sum_i S_{il}$, $S_{i\cdot} = \sum_l S_{il}$ and $S_{\cdot\cdot} = \sum_{i,l} S_{il}$. For each individual, we fitted a smoothing spline to the plot of F_{il} against the mean GC content for the bin. We used the `R` function `smooth.spline` with a smoothing parameter of 1.

Letting $\hat{F}_{il}$ be the predicted value of the smoothing spline for bin l in individual i, we set $c_{ij} = \hat{F}_{il}$, where c_{ij} is the predicted $\log_2$ over/under-representation of fragment (read) count of feature $j \in \mathcal{B}_l$ in individual i. Then the normalised FPKM (RPKM) was obtained by

$$\tilde{y}_{ij} = y_{ij} - c_{ij}.$$

2.5 Principal component correction

There are usually hidden confounding factors, such as sequencing batch, sample preparation date, in real data that reduce power to detect QTLs. These factors are not often observed but can be captured by principal component analysis (PCA) [9]. We applied PCA to log FPKMs with and without permutation and selected the first 16 components for the ATAC-seq, 21 components for gEUVADIS RNA-seq data and 8 components for CTCF ChIP-seq data whose contribution were greater than those obtained from permuted data.

For normalised fragment count data, we regressed out those covariates from the $\log_2$ FPKMs using a standard linear model. Let $\hat{\boldsymbol{\beta}}_j$ be the estimated regression coefficients for feature j and $\boldsymbol{x}_i$ be the vector of covariates for individual i, we use the residual

$$\tilde{y}_{ij} = y_{ij} - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_j$$

for subsequent QTL mapping. Note that we use the ordinary least square method

$$\{\hat{\alpha}_j, \hat{\beta}_j\} = \underset{\{\alpha_j, \beta_j\}}{\operatorname{argmin}} \sum_{i=1}^{N} |y_{ij} - \alpha_j - \boldsymbol{x}_i^\top \beta_j|^2.$$

to estimate coefficients of covariates for each feature j (either gene or peak).

2.6 Genotype imputation

We downloaded VCF files for the 1000 Genomes Phase III integrated variant set from the project website. Because ATAC-seq, RNA-seq and CTCF ChIP-seq samples were not completely overlapping with the 1000 Genomes Phase III samples, we extracted genotype data of the non-overlapping samples from the 1000 Genomes Phase I data and the 1000 Genomes high density SNP chip data (performed on the Illumina Omni platform).

For our ATAC-seq data, HG00104, HG00124, HG00134, HG00135, HG00152, HG00156, HG00247 and HG00249 were extracted from the Phase I data and HG00098 was extracted from the Illumina SNP chip data. For gEUVADIS RNA-seq data, HG00104, HG00124, HG00134, HG00135, HG00152, HG00156, HG00247, HG00249, HG00312, HG00359, HG00377, NA11993, NA20537 and NA20816 were extracted from the Phase I data and NA07346 was extracted from the Illumina SNP chip data. For CTCF ChIP-seq data, NA11993 was extracted from the Phase I data and NA07346, NA12891 and NA12892 were extracted from the Illumina SNP chip data.

We then performed whole genome imputation for the extracted genotype data separately by using the Beagle software (version: 23-Jul-16) [10]. To reduce the computational complexity, we split short and long arms of each chromosome if variants were found in both arms. We finally merged the imputed genotype data with the Phase III data.

2.7 Genomic annotations

Genomic annotations related to chromatin accessibility were generated from our ATAC-seq data. We aggregated all samples' sequencing data into one merged data to generate an average sequencing coverage depth. The peak height was defined as the highest value of the average coverage depth within each peak region (defined in Section 2.2). The peak height was converted into peak height quantiles $\{q_j; j = 1, \ldots, J\} \in (0, 1]$ across all peaks. Combining with VCF information, coverage depth at each genetic variant l (inside the peak a) was normalised with the peak height to compute the relative coverage, $h_{la} \in [0, 1]$. Peak distance was calculated based on the midpoint of a peak region.

We also used various external genomic annotations. The Hi-C contact map, Hi-C loop and TAD information for GM12878 were obtained from [11]. The distance of Hi-C contact domain was calculated from the Hi-C loop list. All data sets were downloaded from the Gene Expression Omnibus (Accession ID: GSE63525). The HiChIP data for GM12878 was obtained from [12]. The JuicerBox output of HiChIP data was downloaded from the Gene Expression Omnibus (Accession ID: GSE101498) and processed by the HiCCUPS software [13] to obtain the HiChIP loop list with the default parameter setting. The capture Hi-C (CHi-C) contact domain information for GM12878 were obtained from [14, 15], from which contact length distribution was generated. The original CHi-C data and annotated interactions for GM12878 [14] was obtained from the ArrayExpress database (Accession ID: E-MTAB-2323). The CHi-C data processed by CHiCAGO pipeline [15] was downloaded from here. The genomic segmentation annotation combining Segway [16] and ChromHMM [17] results was downloaded from the Encode Project. Each ATAC peak was labelled by one of the 7 different segmentation categories at the peak midpoint.

2.8 Roadmap Epigenomics Project data

We downloaded DNaseI-seq data for the 53 different cell types from the project web page. We counted the number of reads mapped at our annotated ATAC peaks and noramlised as FPKM in the same way as our ATAC-seq data. Spearman's correlation was calculated between any peak pair within 500Kb distance across different cell types.

2.9 Bayes factor computation of GWAS summary statistics

We used the GWAS summary statistics for 10 different disease traits, Rheumatoid arthritis (RA), schizophrenia (SCZ), systemic lupus erythematosus (SLE), Crohn's disease (CD), ulcerative colitis (UC), inflammatory bowel diseases (IBD), type 2 diabetes (T2D), Alzheimer's disease (AD), atopic dermatitis (ATD) and coronary artery disease (CAD) available from:

RA http://plaza.umin.ac.jp/ yokada/datasource/software.htm

SCZ http://www.med.unc.edu/pgc/results-and-downloads

SLE http://insidegen.com/insidegen-LUPUS-data.html

IBD/CD/UC https://www.ibdgenetics.org/downloads.html

ATD https://data.bris.ac.uk/data/dataset/28uchsdpmub118uex26ylacqm

AD http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php

T2D http://diagram-consortium.org/downloads.html

CAD http://www.cardiogramplusc4d.org/data-downloads/

For the RA GWAS summary statistics, we used the fixed effect inverse variance meta analysis method to combine test statistics for each variant from the genome-wide SNP chip data across 11 studies (excluding the Illumina Immuno-chip studies). We then use the Wakefield approximation [18] to convert the combined test statistics into Bayes factors. We also applied the Wakefield approximation to the summary statistics for SLE. Here we set the prior variance $W = 0.1$ as described in [19].

3 Pairwise hierarchical model

3.1 Interaction hypotheses

The model aims to estimate the probability that a chromatin accessible peak j ($j = 1, \ldots, J$) affects another peak k ($\neq j$) within a 1 Mb window centred at peak j, such that the maximum distance between peak j and k is 500 Kb). We developed a Bayesian pairwise hierarchical model using overlapping association signals of genetic variants surrounding those peaks. We defined the *cis*-window, $\mathcal{W}_{jk}$, surrounding peak j and k, as the union of two 1 Mb windows centred on each peak and assume that the peak pair falls into one of the following five different hypotheses:

H_0 $(null)$: there are no genetic variants in $\mathcal{W}_{jk}$ that associate with either peak;

$H_{1.1}$ $(single)$: there is one causal variant in $\mathcal{W}_{jk}$ that affects peak j;

$H_{1.2}$ $(single)$: there is one causal variant in $\mathcal{W}_{jk}$ that affects peak k;

H_2 $(linkage)$: there are two independent causal variants in $\mathcal{W}_{jk}$, one of which affects peak j and the other one affects peak k;

H_3 $(pleiotropy)$: there is one causal variant in $\mathcal{W}_{jk}$ that affects both two peaks simultaneously and independently;

$H_{4.1}$ $(causality)$: there is one causal variant in $\mathcal{W}_{jk}$ that affects peak j, and also affects peak k only through its association with peak j;

$H_{4.2}$ $(causality)$: there is one causal variant in $\mathcal{W}_{jk}$ that affects peak k, and also affects peak j only through its association with peak k.

Here the single hypothesis (H_1) and causality hypothesis (H_4) are split into two sub-hypotheses ($H_{1.1}/H_{1.2}$ and $H_{4.1}/H_{4.2}$) because of the asymmetrical feature between peak j and k.

The model employs ATAC-seq from multiple unrelated individuals. Let $Y = (y_1, \ldots, y_J)$ denote a matrix of normalised ATAC-seq read counts (measuring degrees of chromatin accessibility), whose jth element is a vector of the read counts for N individuals at peak j, such that $y_j = (y_{1j}, \ldots, y_{Nj})^\top$. We then construct a finite mixture of regression models to jointly observe read counts for any peak pair under the above hypotheses $\mathcal{H} = \{H_0, H_{1.1}, H_{1.2}, H_2, H_3, H_{4.1}, H_{4.2}\}$. For the peak pair j and k, we model the joint probability of y_j and y_k by a regression model $p(y_j, y_k|h)$ under the hypothesis $h \in \mathcal{H}$, which is weighted by the mixture probability $\Phi_{jk}^{(h)}$ with which the j–k peak pair is potentially classified into hypothesis h. The pairwise likelihood [20] of any peak pair in 500 Kb distance is then defined by

$$L_2(\Phi|Y) = \prod_{\substack{1 \le j < k \le J \\ d(j,k) < 5 \times 10^5}} \sum_{h \in \mathcal{H}} \Phi_{jk}^{(h)} p(y_j, y_k|h),$$

where $\Phi = \{\Phi_{jk}^{(h)}\}$ is the set of mixture probabilities and $d(\cdot, \cdot)$ is the distance function of a peak pair based on the peak mid-points. Note that this likelihood is a special case of the pseudo-likelihood and we assume it preserves necessary and sufficient information on causal inference between peaks. The subsequent sections introduce details of the regression model and mixture probability. Note here that, throughout the manuscript, the notation $\mathcal{W}_{jk}$ is also used as a set of variants or peaks in the window. Therefore we denote a variant l or peak j is in the window as $l \in \mathcal{W}_{jk}$ or $j \in \mathcal{W}_{jk}$. This does not meet the

definition of a "set" in a strict mathematical sense, instead it is context-dependent.

3.2 Regression model $p(y_j, y_k|h)$

The form of regression model depends upon the hypothesis h. We assume it is essentially factorised by two marginal probabilities of observing y_j and y_k in a certain way. Let $p(y_j|\phi)$ denote the marginal probability of observing y_j under the assumption that no genetic variants in $\mathcal{W}_{jk}$ associate with peak j, and $p(y_j|x_l)$ denote the marginal probability of observing y_j under the assumption that single causal variant x_l ($l \in \mathcal{W}_{jk}$) associates with peak j. Here $x_l = (x_{1l}, \ldots, x_{Nl})^\top$ is a vector of genotypes for N individuals at a (biallelic) genetic variant (*i.e.*, SNP, INDEL and CNV). Note that each element $x_{il} \in \{0, 1, 2\}$ is the alternative allele dosage for individual i at variant l. In addition, let $p(y_k|y_j)$ denote the marginal probability of observing y_k under that assumption that peak j affects peak k. Given that we know the causal variant l affects peak j (or k), the regression model is assumed to be factorised as follows:

$$p(y_j, y_k|h, \text{ causal variant(s) known}) = \begin{cases} p(y_j|\phi)p(y_k|\phi) & h = H_0 \\ p(y_j|x_l)p(y_k|\phi) & h = H_{1.1} \\ p(y_j|\phi)p(y_k|x_l) & h = H_{1.2} \\ p(y_j|x_l)p(y_k|x_m) & h = H_2 \\ p(y_j|x_l)p(y_k|x_l) & h = H_3 \\ p(y_j|x_l)p(y_k|y_j) & h = H_{4.1} \\ p(y_k|x_l)p(y_j|y_k) & h = H_{4.2} \end{cases}$$

Note that x_m (in the hypothesis H_2) is a vector of genotypes at the causal variant $m \in \mathcal{W}_{jk}$ for peak k that is not identical to the causal variant l for peak j (*i.e.*, $l \neq m$) .

The causal variants $l, m \in \mathcal{W}_{jk}$ affecting peak j and k are unknown. We therefore introduce a *variant-level* prior probability $\pi_{jk}^{(jl)}$ that variant $l \in \mathcal{W}_{jk}$ is the causal variant for peak j, given that the peak j is a QTL (analogously $\pi_{jk}^{(kl)}$ for peak k). We here assume that there is, at most, one causal variant for peak j, so that $\sum_{l \in \mathcal{W}_{jk}} \pi_{jk}^{(jl)} = 1$. Given that there is a single (unknown) causal variant for peak j, the marginal probability of observing y_j can be written as

$$p(y_j) = \sum_{l \in \mathcal{W}_{jk}} \pi_{jk}^{(jl)} p(y_j|x_l),$$

which leads to

$$p(y_j, y_k|h) = \begin{cases} p(y_j|\phi)p(y_k|\phi) & h = H_0 \\ p(y_j)p(y_k|\phi) & h = H_{1.1} \\ p(y_j|\phi)p(y_k) & h = H_{1.2} \\ \dfrac{1}{R_{jk}} \displaystyle\sum_{l,m \in \mathcal{W}_{jk}, l \neq m} \pi_{jk}^{(jl)} \pi_{jk}^{(km)} p(y_j|x_l)p(y_k|x_m) & h = H_2 \\ \displaystyle\sum_{l \in \mathcal{W}_{jk}} \pi_{jk}^{(*l)} p(y_j|x_l)p(y_k|x_l) & h = H_3 \\ p(y_j)p(y_k|y_j) & h = H_{4.1} \\ p(y_j|y_k)p(y_k) & h = H_{4.2} \end{cases}$$

where $\pi_{jk}^{(km)}$ is the prior probability that the variant m is causal for peak k and $R_{jk} = \sum_{l,m \in \mathcal{W}_{jk}, l \neq m} \pi_{jk}^{(jl)} \pi_{jk}^{(km)}$ is the constant multiplication to exclude the possibility that a single causal variant affects the two peaks

under the hypothesis of linkage (H_2). For the hypothesis of pleiotropy (H_3), we implicitly assume that there is an *anchor* peak "a" in $\mathcal{W}_{jk}$ that affects both peak j and k, whose read count y_a can be integrated out from the joint distribution of observing y_j, y_k and y_a under the assumption of conditional independence

$$\begin{aligned} p(y_j, y_k|H_3) &= \frac{1}{A_{jk}} \sum_{a\in\mathcal{W}_{jk}, a\neq j,k} \int p(y_a)p(y_j, y_k|y_a)dy_a \\ &= \frac{1}{A_{jk}} \sum_{a\in\mathcal{W}_{jk}, a\neq j,k} \sum_{l\in\mathcal{W}_{jk}} \pi_{jk}^{(al)} \int p(y_a|x_l)p(y_j|y_a)p(y_k|y_a)dy_a \\ &= \sum_{l\in\mathcal{W}_{jk}} \left[\frac{1}{A_{jk}} \sum_{a\in\mathcal{W}_{jk}, a\neq j,k} \pi_{jk}^{(al)} \right] p(y_j|x_l)p(y_k|x_l) \\ &\equiv \sum_{l\in\mathcal{W}_{jk}} \pi_{jk}^{(*l)} p(y_j|x_l)p(y_k|x_l), \end{aligned} \tag{1}$$

where A_{jk} is the number of peaks in $\mathcal{W}_{jk}$ other than j and k.

The marginal probabilities $p(y_j|x_l)$ and $p(y_j|\phi)$ are calculated from the effect of each individual genotype x_{il} on the chromatin openness y_{ij} for individual i, following the simple linear regression:

$$y_{ij} = \alpha_{0j}^{(l)} + \alpha_{1j}^{(l)} x_{il} + \varepsilon_{ij}^{(l)}, \tag{2}$$

where $(\alpha_{0j}^{(l)}, \alpha_{1j}^{(l)})$ denotes the coefficients of intercept and slope and $\varepsilon_{ij}^{(l)}$ denotes the residual following the normal distribution. Although, the analytical form of the Bayes factor for the simple linear regression is known [21], we employ the asymptotic Bayes factor [18]

$$\begin{aligned} BF_j^{(l)} &\equiv \frac{p(y_j|x_l)}{p(y_j|\phi)} \\ &\approx \sqrt{1 - r_j^{(l)}} \exp\left\{ \frac{(Z_j^{(l)})^2}{2} r_j^{(l)} \right\} \end{aligned}$$

with $r_j^{(l)} = W/(W + \mathrm{Var}(\hat{\alpha}_{1j}^{(l)}))$, $Z_j^{(l)} = \mathcal{Z}(\hat{\alpha}_{1j}^{(l)}/\mathrm{Var}(\hat{\alpha}_{1j}^{(l)})^{\frac{1}{2}})$ where $\{\hat{\alpha}_{1j}^{(l)}, \mathrm{Var}(\hat{\alpha}_{1j}^{(l)})\}$ is the ordinary least square estimation of the slope and its asymptotic variance. We here introduced the statistic transformation $\mathcal{Z}(\cdot)$ that converts Student's t statistic to standard normal Z statistic (see Appendix A for details).

The motivation for using the Bayes factor approximation is to devise an efficient and robust method to compute the marginal probability $p(y_k|y_j)$. Because of the existence of unknown confounding factors and the measurement error, it is often difficult to estimate the effect size of y_{ik} on y_{ij} with the standard linear regression:

$$y_{ik} = \beta_{0k}^{(j)} + \beta_{1k}^{(j)} y_{ij} + \varepsilon_{ik}^{(j)}. \tag{3}$$

We therefore utilise a Mendelian randomisation (MR) technique [22] to robustly estimate the causal effect $\beta_{1k}^{(j)}$ by using the genetic variant $l \in \mathcal{W}_{jk}$ as the instrumental variable. The Bayes factor of the regression of y_k on y_j using the instrumental variable x_l can be calculated by

$$\begin{aligned} BF_k^{(jl)} &\equiv \frac{p(y_k|y_j)}{p(y_k|\phi)} \\ &\approx \sqrt{1 - r_k^{(jl)}} \exp\left\{ \frac{(Z_k^{(jl)})^2}{2} r_k^{(jl)} \right\} \end{aligned}$$

with $r_k^{(jl)} = W/(W + \mathrm{Var}(\hat{\beta}_{1k}^{(jl)})^2)$ and $Z_k^{(jl)} = \mathcal{Z}(\hat{\beta}_{1k}^{(jl)}/\mathrm{Var}(\hat{\beta}_{1k}^{(jl)})^{\frac{1}{2}})$, where

$$\hat{\beta}_{1k}^{(jl)} = \frac{\mathrm{Cov}(y_k, x_l)}{\mathrm{Cov}(y_j, x_l)}$$

$$\mathrm{Var}(\hat{\beta}_{1k}^{(jl)}) = \frac{\mathrm{Var}(y_k - \hat{\beta}_{1k}^{(jl)} y_j)}{(N-2)\mathrm{Var}(y_j)\mathrm{Cor}(y_j, x_l)^2}$$

is the causal effect and its asymptotic variance of the two stage least square (2SLS) estimate (see Appendix B for details). Note that the estimated causal effect $\hat{\beta}_{1k}^{(jl)}$ is also indexed by $l \in \mathcal{W}_{jk}$, because the 2SLS estimate of the true causal effect $\beta_{1k}^{(j)}$ depends on which instrumental variable is used. Here $\mathrm{Var}(\cdot)$, $\mathrm{Cov}(\cdot,\cdot)$ and $\mathrm{Cor}(\cdot,\cdot)$ indicate sample variance (biased), covariance and correlation, respectively.

For convenience, we rewrite the likelihood with "regional" Bayes factors, defined as the average Bayes factor across all variants in $\mathcal{W}_{jk}$ weighted by the variant prior probabilities, such as

$$RBF_{jk}^{(h)} \equiv \frac{p(y_j, y_k|h)}{p(y_j|\phi)p(y_k|\phi)} = \begin{cases} \sum_{l\in\mathcal{W}_{jk}} \pi_{jk}^{(jl)} BF_j^{(l)} & h = H_{1.1} \\ \sum_{l\in\mathcal{W}_{jk}} \pi_{jk}^{(kl)} BF_k^{(l)} & h = H_{1.2} \\ \frac{1}{R_{jk}} \sum_{l,m\in\mathcal{W}_{jk}, l\neq m} \pi_{jk}^{(jl)} \pi_{jk}^{(km)} BF_j^{(l)} BF_k^{(m)} & h = H_2 \\ \sum_{l\in\mathcal{W}_{jk}} \pi_{jk}^{(*l)} BF_j^{(l)} BF_k^{(l)} & h = H_3 \\ \sum_{l\in\mathcal{W}_{jk}} \pi_{jk}^{(jl)} BF_j^{(l)} BF_k^{(jl)} & h = H_{4.1} \\ \sum_{l\in\mathcal{W}_{jk}} \pi_{jk}^{(kl)} BF_k^{(l)} BF_j^{(kl)} & h = H_{4.2} \end{cases} \tag{4}$$

such as

$$L_2(\Phi, \pi) = \prod_{\substack{1\le j<k\le J \\ d(j,k)<5\times10^5}} p(y_j|\phi)p(y_k|\phi) \left[\Phi_{jk}^{(0)} + \sum_{h\in\mathcal{H}_1} \Phi_{jk}^{(h)} RBF_{jk}^{(h)}\right],$$

where $\pi = \{\pi_{jk}^{(jl)}\}$ and $\mathcal{H}_1 = \{H_0, H_{1.1}, H_{1.2}, H_2, H_3, H_{4.1}, H_{4.2}\}$ is the set of alternative hypotheses.

3.3 Mixture probability $\Phi_{jk}^{(h)}$

Because the hierarchical model relies on the genetic associations of peak j and k to infer causality, the mixture probability $\Phi_{jk}^{(h)}$ depends on the *peak-level* prior probability, $\Pi_j^{(1)}$ ($\Pi_k^{(1)}$) that a peak j (or k) is a QTL. However, when neither variant is a putative QTL, it is essentially impossible to know whether those peaks are interacting or not. In order to model the mixture probability $\Phi_{jk}^{(h)}$ for any peak pair regardless of the fact that a peak is a QTL, we further introduce the *peak pair-level* prior probabilities $\{\Psi_{jk}^{(0)}, \Psi_{jk}^{(1)}, \Psi_{jk}^{(2)}, \Psi_{jk}^{(3)}\}$ with which the j–k peak pair potentially falls into the following interacting categories: (0) no interaction; (1) pleiotropy; (2) causality from peak j to k; or (3) causality from peak k to j. Note that these categories are mutually exclusive, such that $\sum_{i=0}^{3} \Psi_{jk}^{(i)} = 1$.

We assume the peak-pair level prior probability is orthogonal to the peak level prior probability, the hypothesis of non-QTL peak pair (H_0) can be subdivided into the following four hypotheses:

$H_{0.0}$ (*no interaction*) : peak j and k are both non-QTL and not interacting;

$H_{0.1}$ (*pleiotropy*) : peak j and k are pleiotropic, but the anchor peak is non QTL;

$H_{0.2}$ (*causality*) : peak j is causally interacting with peak k, but the peak j is non QTL;

$H_{0.3}$ (*causality*) : peak k is causally interacting with peak j, but the peak k is non QTL;

Let $\mathcal{H}_0 = \{H_{0.0}, H_{0.1}, H_{0.2}, H_{0.3}\}$ denotes a set of the null sub-hypotheses so that all the possible interaction hypotheses is $\mathcal{H} = \mathcal{H}_0 \cup \mathcal{H}_1$. Therefore the mixture probability for each hypothesis is essentially written as a product of peak level and peak-pair level prior probabilities, such that

$$\Phi_{jk}^{(h)} = \begin{cases} \Psi_{jk}^{(0)}\Pi_j^{(0)}\Pi_k^{(0)} & h = H_{0.0} \\ \Psi_{jk}^{(1)}\Pi_*^{(0)} & h = H_{0.1} \\ \Psi_{jk}^{(2)}\Pi_j^{(0)} & h = H_{0.2} \\ \Psi_{jk}^{(3)}\Pi_k^{(0)} & h = H_{0.3} \\ \Psi_{jk}^{(0)}\Pi_j^{(1)}\Pi_k^{(0)} & h = H_{1.1} \\ \Psi_{jk}^{(0)}\Pi_j^{(0)}\Pi_k^{(1)} & h = H_{1.2} \\ \Psi_{jk}^{(0)}\Pi_j^{(1)}\Pi_k^{(1)} & h = H_2 \\ \Psi_{jk}^{(1)}\Pi_*^{(1)} & h = H_3 \\ \Psi_{jk}^{(2)}\Pi_j^{(1)} & h = H_{4.1} \\ \Psi_{jk}^{(3)}\Pi_k^{(1)} & h = H_{4.2} \end{cases} \begin{matrix} \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{matrix} \begin{matrix} \\ \\ \\ \\ \text{Null hypotheses } \mathcal{H}_0 \\ \\ \\ \\ \\ \\ \text{Alternative hypotheses } \mathcal{H}_1 \end{matrix} \tag{5}$$

and, as a consequence of the four null sub-hypotheses, the mixture probability of $h = H_0$ is given by

$$\Phi_{jk}^{(0)} = \sum_{h \in \mathcal{H}_0} \Phi_{jk}^{(h)} = \Psi_{jk}^{(0)}\Pi_j^{(0)}\Pi_k^{(0)} + \Psi_{jk}^{(1)}\Pi_*^{(0)} + \Psi_{jk}^{(2)}\Pi_j^{(0)} + \Psi_{jk}^{(3)}\Pi_k^{(0)},$$

where $\Pi_j^{(0)} = 1 - \Pi_j^{(1)}$ and $\Pi_k^{(0)} = 1 - \Pi_k^{(1)}$. Here $\Pi_*^{(1)}$ denotes the prior probability that an unknown anchor peak is a QTL that affects peak j and k simultaneously, while $\Pi_*^{(0)} = 1 - \Pi_*^{(1)}$ is the probability that the anchor peak is non-QTL. Thus the total of mixture probabilities becomes $\sum_{h \in \mathcal{H}} \Phi_{jk}^{(h)} = 1$.

3.4 Hierarchical structure of prior probabilities

A key feature of the hierarchical model is that the three levels of prior probabilities (variant level, peak level and peak-pair level) are allowed to depend on various genomic annotations.

3.4.1 Variant-level prior

We model the variant level prior probability as a function of three annotations: variant type (SNP, INDEL or CNV), peak status (i.e. whether the variants are located inside or outside the peak region) and, for those variants located within peaks, within-peak location, measured as the read depth at variant l, relative to the maximum read depth in the peak such that:

$$\pi_{jk}^{(jl)} = \frac{e^{\eta_{jl}}}{\sum_{m \in \mathcal{W}_{jk}} e^{\eta_{jm}}}$$

with

$$\eta_{jl} = x^{(l)}_{\text{INDEL}}\lambda_1 + x^{(l)}_{\text{CNV}}\lambda_2 + \begin{cases} f_1(h_{jl};\lambda_3,\lambda_5,\ldots,\lambda_9) & l \text{ is inside the focal peak } j \\ f_1(h_{al};\lambda_4,\lambda_5,\ldots,\lambda_9) & l \text{ is inside a flanking peak } a \in \mathcal{W}_{jk}\ (a \neq j) \\ 0 & \text{otherwise} \end{cases}$$

where $\{\lambda_1,\ldots,\lambda_9\}$ are hyper parameters, each of which indicates the log prior odds relative to a SNP variant in a non-peak region as a baseline. The $0-1$ indicator variables $\{x^{(l)}_{\text{INDEL}}, x^{(l)}_{\text{CNV}}\}$ specify whether variant l is an INDEL or CNV rather than a SNP. The normalised coverage depth $h_{jl} \in [0,1]$ at variant l is relative to the peak height (the maximum coverage) of peak j when the variant l exists inside the peak. We fit a non-linear function $f_1(\cdot)$ which is a cubic spline function with 4 equi-spaced knots, governed by the remaining hyper parameters $\{\lambda_3,\ldots,\lambda_9\}$ (see Appendix C for details). The first parameter of the function, λ_3 or λ_4, controls the intercept of f_1 allowing the log odds to be varied if variant l is in the focal peak j or other peak a ($\neq j$), implying that, except for the constant shift, the function form is identical whenever l is in a peak region.

For a pleiotropic peak pair, the variant level prior probability is averaged across all flanking peaks (Eq. 1). To reduce the computational burden, we approximate this at the linear predictor level as:

$$\pi^{(*l)}_{jk} \approx \frac{e^{\eta_{*l}}}{\sum_{m\in\mathcal{W}_{jk}} e^{\eta_{*m}}}$$

with

$$\eta_{*l} = x^{(l)}_{\text{INDEL}}\lambda_1 + x^{(l)}_{\text{CNV}}\lambda_2 + \begin{cases} f_1(h_{al};\lambda_4,\lambda_5,\ldots,\lambda_9) & l \text{ is inside the peak } a \in \mathcal{W}_{jk} \\ 0 & \text{otherwise} \end{cases}$$

Intuitively, the effect size of anchor peak is the same as that of a flanking peak, because it is essentially unknown which anchor peak is causal to j–k peak pair. The rationale is that

$$\frac{1}{A_{jk}} \sum_{a\in\mathcal{W}_{jk}, a\neq j,k} \eta_{al} \to \eta_{*l}$$

holds true if the number of flanking peak tends to be large enough.

3.4.2 Peak-level prior

The peak-level prior probability is modelled by a non-linear function of the peak height, q_j, the quantile of maximum coverage depth at the peak j compared with all peaks. The model is defined by a logistic regression

$$\log \frac{\Pi^{(1)}_j}{\Pi^{(0)}_j} = f_2(q_j;\gamma_1,\ldots,\gamma_6)$$

with a cubic spline function $f_2(\cdot)$ with 4 equi-spaced knots governed by the 6 hyper-parameters $\{\gamma_1,\ldots,\gamma_6\}$.

3.4.3 Peak-pair-level prior

Our model also enables us to assess effects of various genomic annotations on the probability of observing causal or pleiotropic interactions between peaks. The baseline model of Ψ are non-linear functions

of the distance between peak j and k as multinomial logistic regressions:

$$\log \frac{\Psi_{jk}^{(1)}}{\Psi_{jk}^{(0)}} = f_3(d_{jk}; \delta_1, \ldots, \delta_6)$$
$$\log \frac{\Psi_{jk}^{(2)}}{\Psi_{jk}^{(0)}} = \log \frac{\Psi_{jk}^{(3)}}{\Psi_{jk}^{(0)}} = f_4(d_{jk}; \delta_7, \ldots, \delta_{12})$$

where $f_3(\cdot)$ and $f_4(\cdot)$ are cubic spline functions with 4 knots governed by the hyper-parameters $\{\delta_1, \ldots, \delta_6\}$ and $\{\delta_7, \ldots, \delta_{12}\}$, respectively.

3.5 Likelihood and parameter estimation

The pairwise likelihood should be maximised with respect to all parameters simultaneously. However, in our ATAC-seq data for example, there are 277,128 ATAC-seq peaks genome-wide and the number of peak pairs found within the 500 Kb window is 17,349,412. In addition, each *cis*-region $\mathcal{W}_{jk}$ contains more than 2.5 thousand variants on average. It is unrealistic to maximise the likelihood with all parameters at the same time.

Therefore, we split the model parameters into two sets and estimate in stages. Let $\Theta_1 = \{\Pi(\gamma), \pi(\lambda)\}$ be the parameters associated with the marginal structure for each caQTL peak, and $\Theta_2 = \{\Psi(\delta)\}$ be the set of parameters associated with the pairwise structure between caQTL peaks. Our approach involves a two-stage approach which first maximises the marginal likelihood with respect to Θ_1, and then maximises the pairwise likelihood with respect to Θ_2, where Θ_1 is replaced by its estimator $\hat{\Theta}_1$ obtained in the first stage.

Let

$$L_1(\Theta_1) \propto \prod_{1 \le j \le J} \left[\Pi_j^{(0)} + \Pi_j^{(1)} RBF_j \right] \tag{6}$$

be the marginal likelihood with an independence structure temporarily assumed for each peak. Note that this likelihood is identical to the hierarchical model proposed in [23]. Here

$$RBF_j = \sum_{l \in \mathcal{W}_j} \pi_j^{(l)} BF_j^{(l)}$$

is the regional Bayes factor of associations for peak j, averaged across all variants in the cis-window $\mathcal{W}_j$ (500Kb on either side of peak j), weighted by the prior probability

$$\pi_j^{(l)} = \frac{e^{\eta_{jl}}}{\sum_{m \in \mathcal{W}_j} e^{\eta_{jm}}}$$

that l is the causal variant for peak j (*i.e.*, $\sum_{l \in \mathcal{W}_j} \pi_j^{(l)} = 1$). The linear predictor η_{jl} is identical to that in the pairwise likelihood (Section 3.4) to carry on the prior information of causal variant in the first stage into the second stage.

Let

$$L_2(\Theta_2; \hat{\Theta}_1) \propto \prod_{\substack{1 \le j < k \le J \\ d(j,k) < 5 \times 10^5}} \left[\Phi_{jk}^{(0)} + \sum_{h \in \mathcal{H}_1} \Phi_{jk}^{(h)} RBF_{jk}^{(h)} \right] \tag{7}$$

be the pairwise likelihood as a product of finite mixture probabilities over all peak pairs in 500 Kb window. The first-stage estimator $\hat{\Theta}_1$ is fixed in L_2 to reduce the complexity of optimisation. Our definition of pleiotropy implicitly includes an "anchor" peak, in which the putative causal variant is located. The prior probability of an anchor peak being a QTL, $\Pi_*^{(1)}$, was not directly estimated from the marginal likelihood in Eq. 6, and still a free parameter to be optimised. Although there can be an opportunity to estimate $\Pi_*^{(1)}$ in conjunction with Θ_2 in the second stage, we would rather use the moment estimator $\hat{\Pi}_*^{(1)} = \sum_j \hat{\Pi}_j^{(1)} / J$ as a part of $\hat{\Theta}_1$ for simplicity.

Because we use the cubic spline functions to model the prior probabilities, it is necessary to introduce an appropriate penalty to avoid overfitting. To this end, the integrated square of the second derivative is used for each f_i was not smooth (too wiggly, see Appendix C for details). The penalised log likelihoods are now given by:

$$l_1(\Theta_1) = \log L_1(\Theta_1) - \sum_{i=1}^{2} \frac{\nu_i}{2} \int_0^1 f_i''(x)^2 dx$$

$$l_2(\Theta_2; \hat{\Theta}_1) = \log L_2(\Theta_2; \hat{\Theta}_1) - \sum_{i=3}^{4} \frac{\nu_i}{2} \int_0^1 f_i''(x)^2 dx$$

where ν_i $(i = 1, \ldots, 4)$ are the smoothing parameters. We set $\nu_1 = 5.0$, $\nu_2 = \nu_3 = 1.0$.

We used the Expectation-Maximisation (EM) to optimise the likelihoods [24]. In the E-step, we calculate the posterior probability that a peak is a caQTL

$$\overline{Z}_j = \frac{\hat{\Pi}_j^{(1)} R\hat{B}F_j}{\hat{\Pi}_j^{(0)} + \hat{\Pi}_j^{(1)} R\hat{B}F_j} \tag{8}$$

obtained from the marginal likelihood in Eq. 6. Similarly, the posterior probability that the $j-k$ peak pair belongs to each of the interaction categories is obtained by

$$\overline{Z}_{jk}^{(h)} = \begin{cases} \dfrac{\hat{\Phi}_{jk}^{(0)}}{\hat{\Phi}_{jk}^{(0)} + \sum_{i \in \mathcal{H}_1} \hat{\Phi}_{jk}^{(i)} R\hat{B}F_{jk}^{(i)}} & h = H_0 \\ \dfrac{\hat{\Phi}_{jk}^{(h)} R\hat{B}F_{jk}^{(h)}}{\hat{\Phi}_{jk}^{(0)} + \sum_{i \in \mathcal{H}_1} \hat{\Phi}_{jk}^{(i)} R\hat{B}F_{jk}^{(i)}} & h \in \mathcal{H}_1 \end{cases} \tag{9}$$

from the pairwise likelihood in Eq. 7. In the M-step, the penalised iteratively reweighted least squares (P-IRLS) is used to maximise the marginal and pairwise likelihoods with respect to Θ_1 and Θ_2. The standard error of the model parameter is obtained from the approximated Fisher information matrix. See Appendix D and E for details of parameter estimation for the marginal and pairwise likelihood.

Note that, all subsequent analyses are performed based on $\overline{Z}_{jk}^{(h)}$ in Eq. 9. In the main text, posterior probability of causality (PPC) from j to k is denoted by $\text{PPC}_{jk} (= \overline{Z}_{jk}^{(4.1)})$ and the converse is denoted by $\text{PPC}_{kj} (= \overline{Z}_{jk}^{(4.2)})$ so that $\text{PPC} = \overline{Z}_{jk}^{(4.1)} + \overline{Z}_{jk}^{(4.2)}$.

3.6 Finding directed acyclic graphs (DAGs)

Although the result of the pairwise hierarchical model only provides a relationship between any given two peaks, we can use the posterior probability of causality defined in Eq. 9 to construct multiway

interactions of peaks as directed acyclic graphs (DAGs). We first considered the set of all peak pairs as one large bi-directional graph with 277,128 nodes and more than 17 million edges. Then we selected those edges with $\overline{Z}_{jk}^{(4.1)}$ or $\overline{Z}_{jk}^{(4.2)}$ greater than 0.5 to produce independent bi-directional graphs, each of which consists of peaks connected by at least one or more directional edge(s) with $\overline{Z}_{jk}^{(4.1)} > 0.5$ or $\overline{Z}_{jk}^{(4.2)} > 0.5$ (Fig. S8A). Then we came up with a naive algorithm to find the directed acyclic graph (DAG) embedded in those peaks. Firstly, the most likely parent (MLP) is sequentially assigned for each peak and then the cyclic graph is solved by discarding one of the edges with the lowest $\overline{Z}_{jk}^{(4.1)}$ (or $\overline{Z}_{jk}^{(4.2)}$) to find the DAG (Fig. S8B).

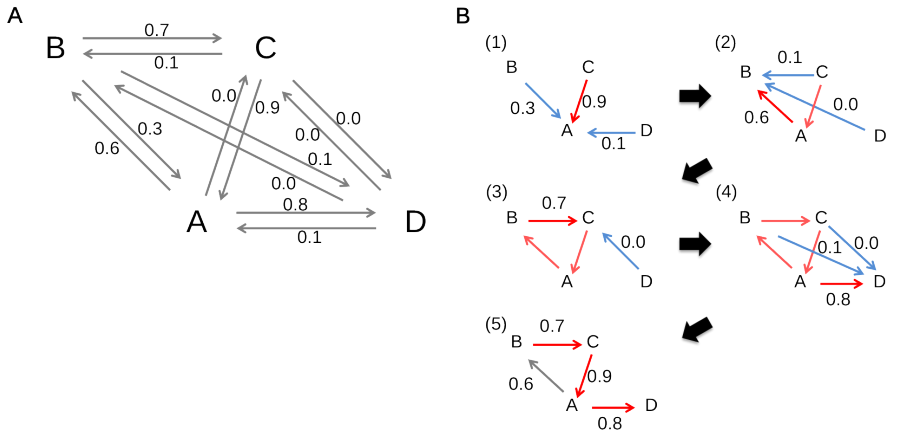


Fig. S8: (A) Bidirected graph shows there are 4 chromatin accessibility peaks (A-D) with arrows depicting direction and posterior probability of causality ($\overline{Z}_{jk}^{(4.1)}$ and $\overline{Z}_{jk}^{(4.2)}$). For example PPC from peak A to D is 80%. (B) The algorithm to find the directed acyclic graph embedded in those peaks: (1) assign the most likely parent (MLP) of A as C (2) assign MLP of B as A (3) assign MLP of C as B (4) assign MLP of D as A (5) discard the least likely arrow from A to B to find the most likely DAG (B→C→A→D).

3.7 Detection of lead caQTL variant

In this section, we describe the procedure we used to identify the most likely putative causal variant for each peak j, using pairwise information from all peaks surrounding this peak. Within each window $\mathcal{W}_j$ (500 Kb on either side of peak j), we calculate a posterior probability of each variant $l \in \mathcal{W}_j$ being the causal caQTL for peak j and obtain the maximum a posteriori (MAP) probability of the lead variant. The marginal likelihood in Eq. 6 naively provides the posterior probability that a variant l is the lead QTL variant, such that:

$$\overline{z}_j^{(l)} = \frac{\hat{\pi}_j^{(l)} BF_j^{(l)}}{\sum_{m \in \mathcal{W}_j} \hat{\pi}_j^{(m)} BF_j^{(m)}}$$

given that the peak j is a QTL. If the peak j is an isolated caQTL (*i.e.* does not interact with any other peak), it is reasonable to select the MAP variant as the lead QTL variant. However it is not appropriate to use the posterior probability when multiple variants in strong linkage disequilibrium are found inside multiple interacting peaks. The reason for this is that $\pi_j^{(l)}$ aggressively upweights variants inside the

peak j as causal variants. We observed that this has the effect that the variant located within the peak j (in LD with the true causal variant) will always become the lead variant, even if the peak j is a downstream peak in a regulatory cascade.

Therefore, we utilise the pairwise likelihood in Eq. 7 to define the lead caQTL variant for each peak j by solving causal interactions between peaks. Let $z_j^{(l)}$ be the indicator variable which is 1 if variant l is causal (regardless of whether the causality is directly affecting peak j or mediated by any other peaks), otherwise 0. The partial likelihood around the peak j is defined as

$$L_2(\hat{\Theta}|y_k; k \in \mathcal{W}_j) = \sum_{k \in \mathcal{W}_j} \sum_{l \in \mathcal{W}_j} \sum_{h \in \mathcal{H}_j} \frac{1}{A_j} p(y_j, y_k | z_j^{(l)}, h_{jk}) p(z_j^{(l)} | h_{jk}) \hat{\Phi}_{jk}^{(h)},$$

where $\mathcal{H}_j = \{H_{1.1}, H_2, H_3, H_{4.1}, H_{4.2}\}$ denotes the set of interaction hypotheses given that the peak j is a caQTL and A_j denotes the number of peaks in $\mathcal{W}_j$ other than the peak j. The regression model to jointly observe y_j and y_k given that l is the causal variant is defined in Section 3.2, such that

$$p(y_j, y_k | z_j^{(l)}, h_{jk}) \propto \begin{cases} BF_j^{(l)} & h_{jk} = H_{1.1} \\ BF_j^{(l)} \dfrac{1}{R_{jk}} \displaystyle\sum_{m \in \mathcal{W}_j, m \neq l} \hat{\pi}_{jk}^{(km)} BF_k^{(m)} & h_{jk} = H_2 \\ BF_j^{(l)} BF_k^{(l)} & h_{jk} = H_3 \\ BF_j^{(l)} BF_k^{(jl)} & h_{jk} = H_{4.1} \\ BF_k^{(l)} BF_j^{(kl)} & h_{jk} = H_{4.2} \end{cases}$$

except for the linkage hypothesis (H_2), where associations to peak k is marginalised across all variants m ($m \in \mathcal{W}_{jk}$, $m \neq l$) because it is not of interest which variant m is causal to peak k. The prior probability that the variant l is causal is also the same as before,

$$p(z_j^{(l)} | h_{jk}) = \begin{cases} \hat{\pi}_{jk}^{(kl)} & h_{jk} = H_{4.2} \\ \hat{\pi}_{jk}^{(jl)} & \text{otherwise} \end{cases}$$

Here we assume that the caQTL effect of variant l is mediated by peak k in $H_{4.2}$, otherwise the variant l directly affects peak j.

Hence, the posterior probability that the variant l is causal to peak j is obtained by Bayes' rule, such as

$$\bar{z}_j^{(l)} = \frac{\sum_{k \in \mathcal{W}_j} z_{jk}^{(jl)}}{\sum_{k \in \mathcal{W}_j} \sum_{m \in \mathcal{W}_j} z_{jk}^{(jm)}},$$

where

$$\begin{aligned} z_{jk}^{(jl)} &= \sum_{h \in \mathcal{H}_j} p(y_j, y_k | z_j^{(l)}, h_{jk}) p(z_j^{(l)} | h_{jk}) \hat{\Phi}_{jk}^{(h)} \\ &= \hat{\Psi}_{jk}^{(0)} [\hat{\Pi}_j^{(1)} \hat{\pi}_{jk}^{(jl)} BF_j^{(l)}] \left[\hat{\Pi}_k^{(0)} + \frac{1}{R_{jk}} \hat{\Pi}_k^{(1)} \sum_{m \in \mathcal{W}_{jk}, m \neq l} \hat{\pi}_{jk}^{(km)} BF_k^{(m)} \right] \\ &\quad + \hat{\Psi}_{jk}^{(1)} \hat{\Pi}_*^{(1)} \hat{\pi}_{jk}^{(*l)} BF_j^{(l)} BF_k^{(l)} + \hat{\Psi}_{jk}^{(2)} \hat{\Pi}_j^{(1)} \hat{\pi}_{jk}^{(jl)} BF_j^{(l)} BF_k^{(jl)} + \hat{\Psi}_{jk}^{(3)} \hat{\Pi}_k^{(1)} \hat{\pi}_{jk}^{(kl)} BF_k^{(l)} BF_j^{(kl)} \end{aligned}$$

denotes the strength of association of the variant l to the peak j, which is either direct or mediated through peak k. Note that we implicitly assume $p(y_k|\phi) = p(y_{k'}|\phi)$ for $k \neq k'$, which holds true when y_k's are normalised so that $\mathrm{Var}(y_k) = \mathrm{Var}(y_{k'})$.

3.8 Detection of lead caQTL variant(s) for a peak pair

The lead variant for each $j-k$ peak pair under the causality hypothesis is defined as the maximum a posteriori variant, such that

$$l = \underset{l \in \mathcal{W}_{jk}}{\text{argmax}}\ \hat{\pi}_{jk}^{(jl)} BF_j^{(l)} BF_k^{(jl)}.$$

Likewise, the lead variant under the reverse causality is also defined as

$$m = \underset{m \in \mathcal{W}_{jk}}{\text{argmax}}\ \hat{\pi}_{jk}^{(km)} BF_k^{(m)} BF_j^{(km)}.$$

Under the pleiotropic hypothesis, the lead variant is defined as

$$l = \underset{l \in \mathcal{W}_{jk}}{\text{argmax}}\ \hat{\pi}_{jk}^{(*l)} BF_j^{(l)} BF_k^{(l)}.$$

For the linkage hypothesis, there are two causal variants l and m that are defined as

$$(l, m) = \underset{l,m \in \mathcal{W}_{jk}, l \neq m}{\text{argmax}}\ \hat{\pi}_{jk}^{(jl)} \hat{\pi}_{jk}^{(km)} BF_j^{(l)} BF_k^{(m)}.$$

3.9 Probability of master regulator (PMR)

We here define the probability that a caQTL peak j is the master regulatory peak. Firstly, by using the posterior probabilities of interaction hypotheses $\{\overline{Z}_{jk}^{(h)}\}$ in Eq. 9, we denote the upstream/downstream probability :

$$\begin{aligned} p(j \text{ is upstream of } k | j \text{ is caQTL}) &= \begin{cases} \dfrac{\overline{Z}_{jk}^{(4.1)}}{\overline{Z}_{jk}^{(1.1)} + \overline{Z}_{jk}^{(2)} + \overline{Z}_{jk}^{(3)} + \overline{Z}_{jk}^{(4.1)} + \overline{Z}_{jk}^{(4.2)}} & j < k \\ \dfrac{\overline{Z}_{kj}^{(4.2)}}{\overline{Z}_{kj}^{(1.2)} + \overline{Z}_{kj}^{(2)} + \overline{Z}_{kj}^{(3)} + \overline{Z}_{kj}^{(4.1)} + \overline{Z}_{kj}^{(4.2)}} & k < j \end{cases} \\ &\equiv p_{j \to k}^{(j)} \\ p(j \text{ is downstream of } k \text{ or } a | j \text{ is caQTL}) &= \begin{cases} \dfrac{\overline{Z}_{jk}^{(3)} + \overline{Z}_{jk}^{(4.2)}}{\overline{Z}_{jk}^{(1.1)} + \overline{Z}_{jk}^{(2)} + \overline{Z}_{jk}^{(3)} + \overline{Z}_{jk}^{(4.1)} + \overline{Z}_{jk}^{(4.2)}} & j < k \\ \dfrac{\overline{Z}_{kj}^{(3)} + \overline{Z}_{kj}^{(4.1)}}{\overline{Z}_{kj}^{(1.2)} + \overline{Z}_{kj}^{(2)} + \overline{Z}_{kj}^{(3)} + \overline{Z}_{kj}^{(4.1)} + \overline{Z}_{kj}^{(4.2)}} & k < j \end{cases} \\ &\equiv p_{a,k \to j}^{(j)} \end{aligned}$$

Note that the peak a is the hypothesised anchor peak that affects both peak j and k in a pleiotropic fashion, thereby it is upstream of the peak j. To use these notations, we define

$$
\begin{aligned}
&p(j \text{ is master regulatory peak}) \\
&\quad = p(j \text{ is master regulatory peak} | j \text{ is caQTL}) p(j \text{ is caQTL}) \\
&\quad = p(j \text{ has downstream peaks} | j \text{ is caQTL}) p(j \text{ has no upstream peak} | j \text{ is caQTL}) \overline{Z}_j \\
&\quad = \left[p(j \text{ has no downstream peak} | j \text{ is caQTL}) \right] p(j \text{ has no upstream peak} | j \text{ is caQTL}) \overline{Z}_j \\
&\quad = \left[1 - \prod_{k \in \mathcal{W}_j} \left(1 - p^{(j)}_{j \to k} \right) \right] \prod_{k \in \mathcal{W}_j} \left(1 - p^{(j)}_{a,k \to j} \right) \overline{Z}_j \\
&\quad \equiv \mathcal{P}_j,
\end{aligned}
$$

where $\overline{Z}_j$ is the posterior probability that peak j is a caQTL in Eq. 8.

3.10 Hierarchical model for expression QTL mapping

We use the standard hierarchical model [23] to perform fine-mapping of putative causal eQTL variants with various genomic annotations. We used RNA-seq data from LCLs in 372 European samples (gEUVADIS Project [4]). For each gene k, we computed the Bayes factors, $BF^{(l)}_k$, for all variants l existing in the cis-regulatory window $\mathcal{W}_k$ (500 Kb on either side of the transcription start site). Let

$$
L(\Theta) \propto \prod_{k=1}^{K} \left[(1 - \Pi^{(1)}_k) + \Pi^{(1)}_k \sum_{l \in \mathcal{W}_k} \pi^{(l)}_k BF^{(l)}_k \right] \tag{10}
$$

be the likelihood of the hierarchical model with a set of parameter $\Theta = \{\Pi, \pi\}$, where $\Pi^{(1)}_k$ is the prior probability that the gene k is an eQTL, which is modelled as a function of the expression level quantile $q_k \in (0, 1]$ for all genes ($K = 48,325$), such that

$$
\log \frac{\Pi^{(1)}_k}{1 - \Pi^{(1)}_k} = f(q_k; \gamma)
$$

for $k = 1, \ldots, K$, where $f(\cdot)$ is the cubic spline with 6 basis functions weighted by coefficients $\gamma = (\gamma_1, \ldots, \gamma_6)^\top$ (see Appendix C for details). Note that q_k is based on the average CPM value (counts per million) across all individuals.

The prior probability of a variant being the causal eQTL, $\pi^{(l)}_k$, is modelled as a function of overlapping genomic annotations. In order to demonstrate our mapped causal information can improve fine-mapping, we compared various combinations of the following five annotations: (1) inside or outside an ATAC peak (referred to as ATAC); (2) eQTL variant location, relative to an ATAC peak (variant location; VL); (3) promoter capture Hi-C contacts (CHi-C); (4) HiChIP loops from baited promoter regions (HiChIP); and (5) PMR value at each ATAC peak (defined in the previous section). Because HiChIP loops link both enhancers and promoters, we simply used only the HiChIP loops whose anchor region(s) overlaps with a CHi-C baited promoter region to assign a specific gene to each loop. If two anchors are overlapping with different genes, we used the loop twice for each gene. The prior probability $\pi^{(l)}_k$ and

these annotations are connected by the softmax function, such that

$$\pi_k^{(l)} = \begin{cases} \dfrac{1}{\#\text{variants in } \mathcal{W}_k} & \text{Flat} \\ \dfrac{e^{b_{kl}\lambda}}{\sum_{m\in\mathcal{W}_k} e^{b_{km}\lambda}} & \text{HiChIP} \\ \dfrac{e^{a_l b_{kl}\lambda}}{\sum_{m\in\mathcal{W}_k} e^{a_m b_{km}\lambda}} & \text{HiChIP+ATAC} \\ \dfrac{e^{c_{kl}\lambda}}{\sum_{m\in\mathcal{W}_k} e^{c_{km}\lambda}} & \text{CHi-C} \\ \dfrac{e^{a_l\lambda}}{\sum_{m\in\mathcal{W}_k} e^{a_m\lambda}} & \text{ATAC} \\ \dfrac{e^{\eta_l}}{\sum_{m\in\mathcal{W}_k} e^{\eta_m}} & \text{ATAC+VL} \\ \dfrac{e^{a_l c_{kl}\lambda}}{\sum_{m\in\mathcal{W}_k} e^{a_m c_{km}\lambda}} & \text{ATAC+CHi-C} \\ \dfrac{w_l^{\phi} e^{a_l\lambda}}{\sum_{m\in\mathcal{W}_k} w_m^{\phi} e^{a_m\lambda}} & \text{ATAC+PMR} \\ \dfrac{w_l^{\phi} e^{\eta_l}}{\sum_{m\in\mathcal{W}_k} w_m^{\phi} e^{\eta_m}} & \text{ATAC+PMR+VL} \end{cases}$$

where

$$a_l = \begin{cases} 1 & \text{variant } l \text{ is inside an ATAC peak} \\ 0 & \text{otherwise} \end{cases}$$

$$b_{kl} = \begin{cases} 1 & \text{variant } l \text{ is inside a HiChIP loop anchor region for gene } k \\ 0 & \text{otherwise} \end{cases}$$

$$c_{kl} = \begin{cases} 1 & \text{variant } l \text{ is inside a CHi-C baited promoter or enhancer region of gene } k \\ 0 & \text{otherwise} \end{cases}$$

$$\eta_l = \begin{cases} g(h_{jl};\lambda_1,\ldots,\lambda_6) & \text{variant } l \text{ is inside the peak } j \in \mathcal{W}_k \\ 0 & \text{otherwise} \end{cases}$$

$$w_l = \begin{cases} \mathcal{P}_j & \text{variant } l \text{ is inside the ATAC peak } j \in \mathcal{W}_k \\ 1 & \text{otherwise} \end{cases}$$

Here $g(\cdot)$ is the cubic spline with 6 basis functions weighted by coefficients $(\lambda_1,\ldots,\lambda_6)$ (see Appendix C for details) and h_{jl} is the normalised coverage depth at variant l relative to the height of peak j introduced in Section 3.4. The model parameters γ, λ and ϕ were estimated using the EM-algorithm with P-IRLS (see Appendix F for details). Note that, the non-informative prior (referred to as Flat) is independent of any genomic annotation so that each variant in $\mathcal{W}_k$ has the same prior probability to be causal.

The fine-mapping is carried out by means of the posterior probability that the variant $l \in \mathcal{W}_k$ being the putative causal eQTL variant,

$$\bar{z}_k^{(l)} = \frac{\hat{\pi}_k^{(l)} BF_k^{(l)}}{\sum_{m\in\mathcal{W}_k} \hat{\pi}_k^{(m)} BF_k^{(m)}},$$

given that the gene k is an eQTL. We defined the minimum set of loci that account for more than 90% of cumulative posterior probability as the credible set of causal variants for each gene k.

For the fine-mapping of *BLK* locus, we developed the full hierarchical model of eQTL with all the genomic annotations previously used in the pairwise hierarchical model of caQTL. The prior probability is defined as

$$\pi_k^{(l)} = \frac{w_{kl}^{\phi} e^{\eta_l}}{\sum_{m \in \mathcal{W}_k} w_{km}^{\phi} e^{\eta_m}},$$

$$w_{kl} = \begin{cases} \mathcal{P}_j \mathcal{C}_{jk} & \text{variant } l \text{ is inside the ATAC peak } j \in \mathcal{W}_k \\ 1 & \text{otherwise} \end{cases}$$

$$\eta_l = x_{\text{INDEL}}^{(l)} \lambda_1 + x_{\text{CNV}}^{(l)} \lambda_2 + \begin{cases} g(h_{kl}; \lambda_3, \lambda_4, \ldots, \lambda_8) & l \text{ is inside the peak } j \in \mathcal{W}_k \\ 0 & \text{otherwise} \end{cases}$$

All model parameters $\Theta = \{\gamma, \lambda, \phi\}$ are estimated using P-IRLS (see Appendix F for details). Here we restrict our attention into variants that are in a peak potentially colocalised with the gene k, that is, $w_{kl} = \mathcal{C}_{jk}\mathcal{P}_j$ if the variant l overlaps with peak j, where $\mathcal{C}_{jk}$ is the probability of pleiotropy over linkage between gene k and peak j defined in Eq. 11 in Section 3.11. One may notice that $\mathcal{C}_{jk}$ was calculated from the same expression data with which we perform the fine-mapping, and therefore the fine-mapping result will be biased. However, by definition, $\mathcal{C}_{jk}$ is independent of the fact that the gene k is an eQTL or not. Therefore it should be a proper annotation to perform the fine-mapping.

3.11 Colocalisation with other cellular QTLs

The pairwise hierarchical model can be utilised to perform the colocalisation analysis [25] of caQTLs with other cellular QTLs such as expression QTLs (eQTLs). The reduced pairwise hierarchical model without causal interaction hypothesis ($H_{4.1}$ and $H_{4.2}$) allows us to distinguish whether two traits are pleiotropic or independent QTLs in linkage.

Because we don't aim for fine-mapping with this model, we here used the non-informative prior probability on the variant level. Then the regional Bayes factors are independent of any hyper-parameters and thereby constant, such that

$$RBF_{jk}^{(h)} \equiv \frac{1}{L_{jk}} \begin{cases} \sum_{l \in \mathcal{W}_{jk}} BF_j^{(l)} & h = H_{1.1} \\ \sum_{l \in \mathcal{W}_{jk}} BF_k^{(l)} & h = H_{1.2} \\ \sum_{l,m \in \mathcal{W}_{jk}, l \neq m} \frac{1}{(L_{jk}-1)} BF_j^{(l)} BF_k^{(m)} & h = H_2 \\ \sum_{l \in \mathcal{W}_{jk}} BF_j^{(l)} BF_k^{(l)} & h = H_3 \end{cases}$$

where $BF_j^{(l)}$ and $BF_k^{(m)}$ are the Bayes factors for trait j and k at the putative causal variants l and m obtained from the simple linear regressions, respectively. Here L_{jk} is the number of variants in the cis-regulatory window $\mathcal{W}_{jk}$ so that the prior probability that variant l is causal is $1/L_{jk}$ in the window. Here $\mathcal{W}_{jk}$ is defined by the union of two 1 Mb cis-regulatory window for the trait j and k. For example, trait j is chromatin accessibility at peak j and trait k is gene expression for gene k, $\mathcal{W}_{jk}$ is defined by the union of 1 Mb window centred at peak j and the 1 Mb window (500 Kb on either side from the transcription start site) for gene k. We also considered only j–k trait pairs whose distance is 500 Kb or

less for simplicity. For example, if there are J chromatin accessibility peaks and gene expressions for K genes genome-wide, there are potentially $J \times K$ trait pairs between chromatin accessibility and gene expression, but the number of pairs is greatly reduced when we consider only traits that are 500 Kb distant or less.

Using the regional Bayes factor definition, the pairwise likelihood can be written as

$$L_2(\Phi_{jk}^{(h)}) \propto \prod_{\substack{1\le j\le J \\ 1\le k\le K \\ d(j,k)<5\times 10^5}} \left[\Phi_{jk}^{(0)} + \sum_{h\in\mathcal{H}_1} \Phi_{jk}^{(h)} RBF_{jk}^{(h)}\right],$$

where $\mathcal{H}_1 = \{H_{1.1}, H_{1.2}, H_2, H_3\}$. The mixture probability $\Phi_{jk}^{(h)}$ for the j–k trait pair under the hypothesis h can be a function of various genomic annotations as before (such as genomic distance between trait j and k). However, because we are not aiming to learn the prior probabilities here, we assume it is constant for any trait pair, such that

$$\Phi_{jk}^{(h)} = \begin{cases} \Psi^{(0)}\Pi^{(0)}\Xi^{(0)} + \Psi^{(1)}\Delta^{(0)} & h = H_0 \\ \Psi^{(0)}\Pi^{(1)}\Xi^{(0)} & h = H_{1.1} \\ \Psi^{(0)}\Pi^{(0)}\Xi^{(1)} & h = H_{1.2} \\ \Psi^{(0)}\Pi^{(1)}\Xi^{(1)} & h = H_2 \\ \Psi^{(1)}\Delta^{(1)} & h = H_3 \end{cases}$$

where $\Psi^{(1)}$ denotes the prior probability that a trait pair is pleiotropic, $\Pi^{(1)}$ denotes the prior probability that the trait j is a QTL, $\Xi^{(1)}$ denotes the prior probability that the trait k is a QTL, $\Delta^{(1)}$ denotes the prior probability that an anchor trait (affecting the trait j and k simultaneously) is a QTL, and $\Psi^{(0)} = 1 - \Psi^{(1)}$, $\Pi^{(0)} = 1 - \Pi^{(1)}$, $\Xi^{(0)} = 1 - \Xi^{(1)}$, and $\Delta^{(0)} = 1 - \Delta^{(1)}$. Note that the anchor trait can also be either trait j or k itself, if there is causality between them. The maximum likelihood estimator of these parameters $\{\Pi^{(1)}, \Psi^{(1)}, \Xi^{(1)}, \Delta^{(1)}\}$ can be easily obtained by a standard EM-algorithm (see Appendix G for details).

In analogy with the pairwise hierarchical model for the single trait, the posterior probability $\overline{Z}_{jk}^{(h)}$ can be obtained from Eq. 9. The number of expected colocalised traits for the trait k (such as the number of ATAC peaks that colocalised with the gene k) can be calculated by the sum of posterior probabilities

$$\mathcal{N}_k \equiv \sum_{j\in\mathcal{W}_k} \overline{Z}_{jk}^{(h)},$$

where $\mathcal{W}_k$ is the cis-regulatory window for trait k (*e.g.*, 500 Kb on either side of the transcription start site of gene k). The set of genes with $\mathcal{N}_k > 1$ for $k = 1, \ldots, K$ ($K = 48,325$) was used to assess the efficiency of fine-mapping eQTLs (Fig. 4A in the main text). The posterior probability of pleiotropy over linkage is defined by

$$\mathcal{C}_{jk} \equiv \frac{\overline{Z}_{jk}^{(3)}}{\overline{Z}_{jk}^{(2)} + \overline{Z}_{jk}^{(3)}} \tag{11}$$

which denotes the probability of pleiotropy given that trait j and k are both QTLs. This probability is used in the hierarchical model of fine-mapping eQTLs in the previous section.

By using the colocalisation model, we also calculate the probability that ATAC peaks j and j' are jointly colocalised with an eQTL gene k under the causal interaction mapped between j and j'. However, the colocalisation model assumes each trait j (or k) is independent. Therefore we approximate the

probability by its lower bound:

$$p(j \text{ and } j' \text{ are both colocalised with } k|j \text{ and } j' \text{ are QTLs}) \\ \leq p(j \text{ and } j' \text{ are both colocalised with } k|j \text{ and } j' \text{ are causally interacting}),$$

because the fact that j and j' are both QTLs is a necessary condition of the existence of mapped causal interaction, such that $p(j \text{ and } j' \text{ are QTLs}) \geq p(j \text{ and } j' \text{ are causally interacting})$. Let A_j be the binary random variable, such that $A_j = 1$ if j is a QTL; otherwise $A_j = 0$. Likewise B_k denotes the binary random variable, such that $B_k = 1$ if k is a QTL; otherwise $B_k = 0$. Let C_{jk} denote the binary random variable, such that $C_{jk} = 1$ if traits j and k are colocalised, otherwise $C_{jk} = 0$. Then the lower bound of the colocalisation probability can be obtained by

$$\begin{aligned}
\mathcal{D}_{jj'k} &\equiv p(C_{jk} = 1, C_{j'k} = 1 | A_j = 1, A_{j'} = 1) \\
&= \sum_{b_k=0}^{1} p(C_{jk} = 1, C_{j'k} = 1, B_k = b_k | A_j = 1, A_{j'} = 1) \\
&= \sum_{b_k=0}^{1} p(C_{jk} = 1, C_{j'k} = 1 | B_k = b_k, A_j = 1, A_{j'} = 1) p(B_k = b_k | A_j = 1, A_{j'} = 1) \\
&= \sum_{b_k=0}^{1} p(C_{jk} = 1 | B_k = b_k, A_j = 1) p(C_{jk} = 1 | B_k = b_k, A_{j'} = 1) p(B_k = b_k | A_j = 1, A_{j'} = 1) \\
&= \mathcal{C}_{jk}\mathcal{C}_{j'k} p(B_k = 1 | A_j = 1, A_{j'} = 1) \\
&= \mathcal{C}_{jk}\mathcal{C}_{j'k} \frac{p(A_j = 1, A_{j'} = 1 | B_k = b_k) p(B_k = 1)}{p(A_j = 1, A_{j'} = 1)} \\
&= \mathcal{C}_{jk}\mathcal{C}_{j'k} \frac{p(A_j = 1 | B_k = 1) p(A_{j'} = 1 | B_k = 1) p(B_k = 1)}{p(A_j = 1) p(A_{j'} = 1)} \\
&= \frac{\overline{Z}_{jk}^{(3)}}{(\overline{Z}_{jk}^{(1.1)} + \overline{Z}_{jk}^{(2)} + \overline{Z}_{jk}^{(3)})(\overline{Z}_{jk}^{(1.2)} + \overline{Z}_{jk}^{(2)} + \overline{Z}_{jk}^{(3)})} \frac{\overline{Z}_{j'k}^{(3)}}{(\overline{Z}_{j'k}^{(1.1)} + \overline{Z}_{j'k}^{(2)} + \overline{Z}_{j'k}^{(3)})(\overline{Z}_{j'k}^{(1.2)} + \overline{Z}_{j'k}^{(2)} + \overline{Z}_{j'k}^{(3)})} \overline{Z}_k
\end{aligned}$$

since $p(C_{jk} = 1 | A_j = 1, B_k = 0) = 0$, $p(A_j = 1) = \overline{Z}_{jk}^{(1.1)} + \overline{Z}_{jk}^{(2)} + \overline{Z}_{jk}^{(3)}$ and $p(A_j = 1 | B_k = 1) = (\overline{Z}_{jk}^{(2)} + \overline{Z}_{jk}^{(3)}) / (\overline{Z}_{jk}^{(1.2)} + \overline{Z}_{jk}^{(2)} + \overline{Z}_{jk}^{(3)})$. Here $\overline{Z}_k$ was obtained from Eq. 14 (see Appendix. D) for the eQTL hierarchical model without annotation (referred to as "Flat" in Section 3.10). Using this probability we can calculate the probability that the peak pair j and j' are jointly colocalised with at least one gene

$$\begin{aligned}
D_{jj'} &\equiv p(j \text{ and } j' \text{ jointly colocalised with} \geq 1 \text{ genes} | A_j = 1, A_{j'} = 1) \\
&= 1 - \prod_k (1 - \mathcal{D}_{jj'k}).
\end{aligned} \tag{12}$$

The probability is used to calculate the enrichment of causal interactions for the ENCODE segmentation categories in the next section.

3.12 Colocalisation analysis with GWAS traits

We also used the reduced pairwise hierarchical model (Section 3.11) to colocalise a GWAS locus to a caQTL peak. We used genetic associations for the ATAC peak j and a GWAS trait within the 1Mb cis-window $\mathcal{W}_j$ centred at peak j. The regional Bayes factors for the 4 alternative hypotheses are given

by

$$RBF_j^{(h)} \equiv \frac{1}{L_j} \begin{cases} \sum_{l \in \mathcal{W}_j} BF_j^{(l)} & h = H_{1.1} \\ \sum_{l \in \mathcal{W}_j} BF_{\text{GWAS}}^{(l)} & h = H_{1.2} \\ \sum_{l,m \in \mathcal{W}_j, l \neq m} \frac{1}{(L_j - 1)} BF_j^{(l)} BF_{\text{GWAS}}^{(m)} & h = H_2 \\ \sum_{l \in \mathcal{W}_j} BF_j^{(l)} BF_{\text{GWAS}}^{(l)} & h = H_3 \end{cases}$$

where hypothesis H_3 is colocalisation between a caQTL and a GWAS locus. Here $BF_{\text{GWAS}}^{(l)}$ denotes the Bayes factor of disease association at genetic locus $l \in \mathcal{W}_j$ computed from GWAS summary statistics (see Section 2.9 for details). The pairwise likelihood between chromatin accessibility and a GWAS trait is given by

$$L_2(\Phi_j^{(h)}) \propto \prod_{1 \leq j \leq J} \left[\Phi_j^{(0)} + \sum_{h \in \mathcal{H}_1} \Phi_j^{(h)} RBF_j^{(h)} \right],$$

where $\mathcal{H}_1 = \{H_{1.1}, H_{1.2}, H_2, H_3\}$ denotes a set of alternative hypotheses and

$$\Phi_j^{(h)} = \begin{cases} \Psi^{(0)} \Pi^{(0)} \Xi^{(0)} + \Psi^{(1)} \Delta^{(0)} & h = H_0 \\ \Psi^{(0)} \Pi^{(1)} \Xi^{(0)} & h = H_{1.1} \\ \Psi^{(0)} \Pi^{(0)} \Xi^{(1)} & h = H_{1.2} \\ \Psi^{(0)} \Pi^{(1)} \Xi^{(1)} & h = H_2 \\ \Psi^{(1)} \Delta^{(1)} & h = H_3 \end{cases}$$

is the mixture probability for each hypothesis h.

Because of the limited instances ($j = 1, \ldots, J$) to estimate hyperparameters $\{\Psi, \Pi, \Xi, \Delta\}$, we assumed that the probability Δ_1 (that there exists a genetic variant l affecting chromatin accessibility at peak j and a GWAS trait, simultaneously) is identical to the probability Π_1 (that there exists a genetic variant l affecting chromatin accessibility at peak j) to reduce the degrees of freedom. In addition, $\hat{\Pi}^{(1)}$ and $\hat{\Xi}^{(1)}$ are estimated by maximising the following likelihood functions

$$L_1(\Pi) \propto \prod_{1 \leq j \leq J} \left[\Pi^{(0)} + \Pi^{(1)} RBF_j^{(1.1)} \right],$$
$$L_1(\Xi) \propto \prod_{1 \leq j \leq J} \left[\Xi^{(0)} + \Xi^{(1)} RBF_j^{(1.2)} \right]$$

with respect to $\Pi^{(1)}$ and $\Xi^{(1)}$ in advance. We then plugged $\{\hat{\Pi}, \hat{\Xi}, \hat{\Delta}\}$ into the pairwise likelihood to maximise with respect to Ψ.

The posterior probability for each hypothesis was computed from

$$\overline{Z}_j^{(h)} = \begin{cases} \frac{\hat{\Phi}_j^{(0)}}{\hat{\Phi}_j^{(0)} + \sum_{i \in \mathcal{H}_1} \hat{\Phi}_j^{(i)} RBF_j^{(i)}} & h = H_0 \\ \frac{\hat{\Phi}_j^{(h)} RBF_j^{(h)}}{\hat{\Phi}_j^{(0)} + \sum_{i \in \mathcal{H}_1} \hat{\Phi}_j^{(i)} RBF_j^{(i)}} & h \in \mathcal{H}_1 \end{cases}$$

and the probability of colocalisation given peak j is a caQTL is given by

$$p(\text{GWAS Colocalisation}|j \text{ is caQTL}) = \frac{\overline{Z}_j^{(3)}}{\overline{Z}_j^{(1.1)} + \overline{Z}_j^{(2)} + \overline{Z}_j^{(3)}}$$

which is used for the enrichment analysis of GWAS locus given that a ATAC peak is a caQTL.

3.13 Enrichment analysis

Any enrichment analysis is carried out based on the posterior probability of causality (PPC) between peak j and k, $\overline{Z}_{jk}^{(4.1)}$ and $\overline{Z}_{jk}^{(4.2)}$ in Eq. 9. We don't introduce any arbitrary threshold on PPC to binarise the data. For a non-directional annotation X_{jk} for the j–k peak pair (*e.g.*, overlapping with a topologically associating domain)

$$X_{jk} = \begin{cases} 1 & j\text{–}k \text{ peak pair overlaps with a genomic annotation} \\ 0 & \text{otherwise,} \end{cases}$$

we compute the 2×2 table

$$T = \sum_{\substack{1 \le j < k \le J \\ d(j,k) < 5 \times 10^5}} \begin{pmatrix} X_{jk}(\overline{Z}_{jk}^{(4.1)} + \overline{Z}_{jk}^{(4.2)}) & (1 - X_{jk})(\overline{Z}_{jk}^{(4.1)} + \overline{Z}_{jk}^{(4.2)}) \\ X_{jk}(1 - \overline{Z}_{jk}^{(4.1)} - \overline{Z}_{jk}^{(4.2)}) & (1 - X_{jk})(1 - \overline{Z}_{jk}^{(4.1)} - \overline{Z}_{jk}^{(4.2)}) \end{pmatrix}$$

as if $\overline{Z}_{jk}^{(4.1)}$ and $\overline{Z}_{jk}^{(4.2)}$ are binary variables. The hypothesis testing of independence between causal peak pair and the annotation is performed by

$$\log OR = \log \frac{T_{11} T_{22}}{T_{12} T_{21}} \sim \mathcal{N}\left(0, \sqrt{\frac{1}{T_{11}} + \frac{1}{T_{12}} + \frac{1}{T_{21}} + \frac{1}{T_{22}}}\right),$$

where T_{ij} is the i–j element of the 2×2 table T.

We can utilise peak level annotation, such as an ENCODE genome segmentation, to characterise causal peak pairs. Let A_j be a binary variable which is $A_j = 1$ if peak j overlaps with the annotation A, otherwise $A_j = 0$. Likewise, let B_j be another binary variable for the annotation B. We define a directional annotation Y_{jk}/Y_{kj}, such that

$$Y_{jk} = \begin{cases} 1 & \text{peak } j \text{ overlaps with annotation } A \text{ and } k \text{ overlaps with } B \ (A_j = 1 \text{ and } B_k = 1) \\ 0 & \text{otherwise} \end{cases}$$

$$Y_{kj} = \begin{cases} 1 & \text{peak } k \text{ overlaps with annotation } A \text{ and } j \text{ overlaps with } B \ (A_k = 1 \text{ and } B_j = 1) \\ 0 & \text{otherwise} \end{cases}$$

Note that $Y_{jk} \neq Y_{kj}$ in general. For the non-directional annotation, we compute the 2×2 table

$$T^D = \frac{1}{2} \sum_{\substack{1 \le j < k \le J \\ d(j,k) < 5 \times 10^5}} \begin{pmatrix} Y_{jk}\overline{Z}_{jk}^{(4.1)} + Y_{kj}\overline{Z}_{jk}^{(4.2)} & (1 - Y_{jk})\overline{Z}_{jk}^{(4.1)} + (1 - Y_{kj})\overline{Z}_{jk}^{(4.2)} \\ Y_{jk}(1 - \overline{Z}_{jk}^{(4.1)}) + Y_{kj}(1 - \overline{Z}_{jk}^{(4.2)}) & (1 - Y_{jk})(1 - \overline{Z}_{jk}^{(4.1)}) + (1 - Y_{kj})(1 - \overline{Z}_{jk}^{(4.2)}) \end{pmatrix}$$

for the hypothesis testing of independence. The reason of the table divided by 2 is to avoid double counting so that the total number of cells, $T_{11}^D + T_{12}^D + T_{21}^D + T_{22}^D$, is equal to the number of peak pairs used in the analysis.

Intuitively, the odds ratio calculated from the T^D captures the degree of asymmetry S between annotations A and B as well as enrichment E of the annotation pair A and B for the causal peak pair. Let p be the probability that peak j and k are causally interacting (either way) and q be the peak j overlapping with the annotation A and peak k overlapping with the annotation B (or vice versa). The underlying model is written as

	$j = A, k = B$	$j = B, k = A$	otherwise	Total
$j \to k$	$pq + S + E$	$pq - S + E$	$p(1-2q) - 2E$	p
$j \leftarrow k$	$pq - S + E$	$pq + S + E$	$p(1-2q) - 2E$	p
otherwise	$(1-2p)q - 2E$	$(1-2p)q - 2E$	$(1-2p)(1-2q) + 4E$	$1-2p$
Total	q	q	$1-2q$	1

where $S > 0$ suggests the annotation pair $A \to B$ drives the causality (*i.e.*, A is more likely to be upstream peaks and B is more likely to be downstream peaks) and $E > 0$ suggests A–B annotation pair is enriched for the causal peak pairs. Namely, $S = 0$ and $E = 0$ suggest the A–B annotation pair is independent of causality between peaks. It is worth noting that

$$\mathbb{E}[T^D] = (T^D_{11} + T^D_{12} + T^D_{21} + T^D_{22}) \times \begin{pmatrix} pq + S + E & p(1-q) - S - E \\ (1-p)q - S - E & (1-p)(1-q) + S + E \end{pmatrix}$$

suggesting the hypothesis testing on T^D tests the compound effect $S + E = 0$ as the null hypothesis and $S + E > 0$ for the alternative hypothesis.

We use the joint colocalisation probability defined in Eq. 12 to calculate the enrichment of an annotation Y_{jk}/Y_{kj} for the mapped causal interactions that are also jointly colocalised with at least one eQTL. We compute the following 2×2 table

$$T^D = \frac{1}{2} \sum_{\substack{1 \le j < k \le J \\ d(j,k) < 5 \times 10^5}} \begin{pmatrix} Y_{jk}\overline{Z}^{(4.1)}_{jk}\mathcal{D}_{jk} + Y_{kj}\overline{Z}^{(4.2)}_{jk}\mathcal{D}_{jk} & (1-Y_{jk})\overline{Z}^{(4.1)}_{jk}\mathcal{D}_{jk} + (1-Y_{kj})\overline{Z}^{(4.2)}_{jk}\mathcal{D}_{jk} \\ Y_{jk}\overline{Z}^{(4.1)}_{jk}(1-\mathcal{D}_{jk}) + Y_{kj}\overline{Z}^{(4.2)}_{jk}(1-\mathcal{D}_{jk}) & (1-Y_{jk})\overline{Z}^{(4.1)}_{jk}(1-\mathcal{D}_{jk}) + (1-Y_{kj})\overline{Z}^{(4.2)}_{jk}(1-\mathcal{D}_{jk}) \end{pmatrix},$$

suggesting the enrichment is calculated over causally interacting peak pairs that are "not" colocalised with any eQTL.

To adjust a confounding effect, such as the peak height effect, for the enrichment, we stratify all peaks $\mathcal{J} = \{1, \ldots, J\}$ into 10 equally-sized quantile bins $\mathcal{B}_l = \{j; (l-1)/10 < q_j \le l/10\}$ for $l = 1, \ldots, 10$, where q_j is the peak height quantile across all peaks as an example. Then we compose the direct product $\mathcal{C}_n = \mathcal{B}_l \times \mathcal{B}_m$ for $l, m = 1, \ldots, 10$, where $n = (l-1) \times 10 + m$. Now we can stratify peak pairs into 100

quantile bins with which we compute the 2×2 tables for each strata $\mathcal{C}_n$, such that

$$T_n = \sum_{\substack{1 \le j < k \le J \\ d(j,k) < 5 \times 10^5 \\ (j,k) \in \mathcal{C}_n}} \begin{pmatrix} X_{jk}(\overline{Z}_{jk}^{(4.1)} + \overline{Z}_{jk}^{(4.2)}) & (1 - X_{jk})(\overline{Z}_{jk}^{(4.1)} + \overline{Z}_{jk}^{(4.2)}) \\ X_{jk}(1 - \overline{Z}_{jk}^{(4.1)} - \overline{Z}_{jk}^{(4.2)}) & (1 - X_{jk})(1 - \overline{Z}_{jk}^{(4.1)} - \overline{Z}_{jk}^{(4.2)}) \end{pmatrix}$$

$$T_n^D = \sum_{\substack{1 \le j < k \le J \\ d(j,k) < 5 \times 10^5 \\ (j,k) \in \mathcal{C}_n}} \begin{pmatrix} Y_{jk}\overline{Z}_{jk}^{(4.1)} & (1 - Y_{jk})\overline{Z}_{jk}^{(4.1)} \\ Y_{jk}(1 - \overline{Z}_{jk}^{(4.1)}) & (1 - Y_{jk})(1 - \overline{Z}_{jk}^{(4.1)}) \end{pmatrix}$$

$$+ \sum_{\substack{1 \le j < k \le J \\ d(j,k) < 5 \times 10^5 \\ (k,j) \in \mathcal{C}_n}} \begin{pmatrix} Y_{kj}\overline{Z}_{jk}^{(4.2)} & (1 - Y_{kj})\overline{Z}_{jk}^{(4.2)} \\ Y_{kj}(1 - \overline{Z}_{jk}^{(4.2)}) & (1 - Y_{kj})(1 - \overline{Z}_{jk}^{(4.2)}) \end{pmatrix}$$

for $n = 1, \ldots, 100$. Odds ratios derived from T_n or T_n^D are then combined using the inverse variance (fixed-effect) method which is often used in meta-analysis.

3.14 Simulation strategy

In order to simulate realistic data, we used the real genotypes in conjunction with estimated effect sizes and standard errors from simple linear regression model and 2SLS model as well as the variant-level prior probability. To simulate chromatin accessibility at peak j, we first selected one causal variant $l \in \mathcal{W}_{jk}$ according to the estimated variant-level prior probability $\{\hat{\pi}_{jk}^{(jl)}\}$ in the window $\mathcal{W}_{jk}$. Then, using Eq. 2, we simulated the chromatin accessibility at peak j for the individual i as

$$y_{ij} \sim \mathcal{N}(\hat{\alpha}_{1j}^{(l)} x_{il}, (\hat{\sigma}_j^{(l)})^2),$$

where $\hat{\sigma}_j^{(l)}$ denotes the estimated standard deviation for $\varepsilon_{ij}^{(l)}$ in Eq. 2. For the linkage hypothesis, we repeated the procedure for peak k by means of $\{\hat{\pi}_{jk}^{(km)}\}$, $\hat{\alpha}_{1k}^{(m)}$ and $\hat{\sigma}_k^{(m)}$, independently. For the causality hypothesis, using Eq. 3, we further simulated the chromatin accessibility at peak k for the individual i as

$$y_{ik} \sim \mathcal{N}(\hat{\beta}_{1k}^{(jl)} y_{ij}, (\hat{\sigma}_{jk}^{(jl)})^2),$$

where $\hat{\beta}_{1k}^{(jl)} = \hat{\alpha}_{1k}^{(l)} / \hat{\alpha}_{1j}^{(l)}$ and $\hat{\sigma}_{jk}^{(jl)}$ denotes the estimated standard deviation for $\varepsilon_{ik}^{(jl)}$ in Eq 3. For the assumption that peak j and k are pleiotropic, we first picked up the causal variant $l \in \mathcal{W}_{jk}$ according to $\{\hat{\pi}_{jk}^{(*l)}\}$, then simulated

$$y_{ij} \sim \mathcal{N}(\hat{\alpha}_{1j}^{(l)} x_{il}, (\hat{\sigma}_j^{(l)})^2),$$
$$y_{ik} \sim \mathcal{N}(\hat{\alpha}_{1k}^{(l)} x_{il}, (\hat{\sigma}_k^{(l)})^2),$$

simultaneously.

The readers might be keen to see the performance of PHM under more complicated scenarios in which the assumptions of our mixture model are potentially violated. We first simulated two causal variants under the hypotheses of linkage, pleiotropy and causality (top row of Supplementary Fig. 9).

In these cases, we picked up the first causal variant $l \in \mathcal{W}_{jk}$ according to $\{\hat{\pi}_{jk}^{(jl)}\}$ and then the second causal variant $l' \in \mathcal{W}_{jk}$ according to $\{\hat{\pi}_{jk}^{(jl')}/(1-\hat{\pi}_{jk}^{(jl)})\}_{l' \neq l}$. The accessibility is simulated as

$$y_{ij} \sim \mathcal{N}(\hat{\alpha}_{1j}^{(l)} x_{il} + \hat{\alpha}_{1j}^{(l')} x_{il'}, (\hat{\sigma}_j^{(ll')})^2),$$

where we took an average of the estimated variances, $(\hat{\sigma}_j^{(ll')})^2 = [(\hat{\sigma}_j^{(l)})^2 + (\hat{\sigma}_j^{(l')})^2]/2$, for the variants l and l'. We repeated the procedure for the peak k in the linkage hypothesis. For simulating the accessibility at peak k downstream of j, we took an average of the MR effect sizes and the estimated variances, $\hat{\beta}_{1k}^{(jll')} = [\hat{\beta}_{1k}^{(jl)} + \hat{\beta}_{1k}^{(jl')}]/2$ and $(\hat{\sigma}_{jk}^{(jll')})^2 = [(\hat{\sigma}_{jk}^{(jl)})^2 + (\hat{\sigma}_{jk}^{(jl')})^2]/2$, for the variants l and l'. Then we simulated

$$y_{ik} \sim \mathcal{N}(\hat{\beta}_{1k}^{(jll')} y_{ij}, (\hat{\sigma}_{jk}^{(jll')})^2).$$

For the hypothesis of pleiotropy, we picked up the two causal variants l and l' according to $\{\hat{\pi}_{jk}^{(*l)}\}$ and $\{\hat{\pi}_{jk}^{(*l')}/(1-\hat{\pi}_{jk}^{(*l)})\}_{l' \neq l}$. Then we simulated

$$y_{ij} \sim \mathcal{N}(\hat{\alpha}_{1j}^{(l)} x_{il} + \hat{\alpha}_{1j}^{(l')} x_{il'}, (\hat{\sigma}_j^{(ll')})^2),$$
$$y_{ik} \sim \mathcal{N}(\hat{\alpha}_{1k}^{(l)} x_{il} + \hat{\alpha}_{1k}^{(l')} x_{il'}, (\hat{\sigma}_k^{(ll')})^2),$$

simultaneously. We also simulated chromatin accessibility at peak j and k under the hybrid hypothesis (bottom row of Supplementary Fig. 9) in which a pair of linkage, pleiotropy and causality was considered. For the hybrid hypothesis of linkage and pleiotropy, we selected $l \in \mathcal{W}_{jk}$ according to $\{\hat{\pi}_{jk}^{(jl)}\}$ and $m \in \mathcal{W}_{jk}$ according to $\{\hat{\pi}_{jk}^{(km)}\}$, then simulated

$$y_{ij} \sim \mathcal{N}(\hat{\alpha}_{1j}^{(l)} x_{il}, (\hat{\sigma}_j^{(l)})^2),$$
$$y_{ik} \sim \mathcal{N}(\hat{\alpha}_{1k}^{(l)} x_{il} + \hat{\alpha}_{1k}^{(m)} x_{im}, (\hat{\sigma}_k^{(lm)})^2),$$

where $(\hat{\sigma}_k^{(lm)})^2 = [(\hat{\sigma}_k^{(l)})^2 + (\hat{\sigma}_k^{(m)})^2]/2$. For the hybrid hypothesis of linkage and causality, we selected $l \in \mathcal{W}_{jk}$ according to $\{\hat{\pi}_{jk}^{(jl)}\}$ and $m \in \mathcal{W}_{jk}$ according to $\{\hat{\pi}_{jk}^{(km)}\}$, then simulated

$$y_{ij} \sim \mathcal{N}(\hat{\alpha}_{1j}^{(l)} x_{il}, (\hat{\sigma}_j^{(l)})^2),$$
$$y_{ik} \sim \mathcal{N}(\hat{\beta}_{1k}^{(jl)} y_{ij} + \hat{\alpha}_{1k}^{(m)} x_{im}, (\hat{\sigma}_{jk}^{(jlm)})^2),$$

where $(\hat{\sigma}_{jk}^{(jlm)})^2 = [(\hat{\sigma}_{jk}^{(jl)})^2 + (\hat{\sigma}_k^{(m)})^2]/2$. For the hybrid hypothesis of pleiotropy and causality, we simulated

$$y_{ij} \sim \mathcal{N}(\hat{\alpha}_{1j}^{(l)} x_{il}, (\hat{\sigma}_j^{(l)})^2),$$
$$y_{ik} \sim \mathcal{N}(\hat{\beta}_{1k}^{(jl)} y_{ij} + \hat{\alpha}_{1k}^{(l)} x_{il}, (\hat{\sigma}_{jk}^{(jll)})^2),$$

where $(\hat{\sigma}_{jk}^{(jll)})^2 = [(\hat{\sigma}_{jk}^{(jl)})^2 + (\hat{\sigma}_k^{(l)})^2]/2$.

3.15 Software

The software is available from GitHub https://github.com/natsuhiko/PHM.

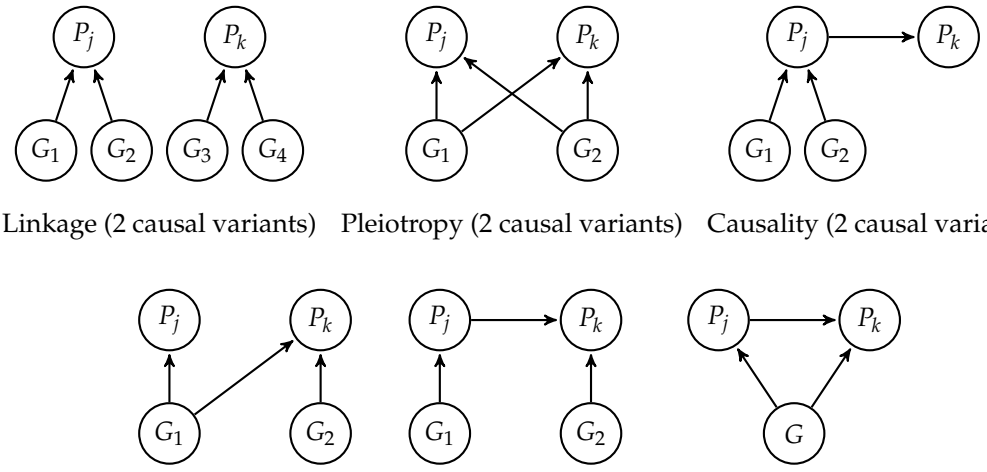


Fig. S9: Schematic of simulation setting. All DAGs above are not considered in PHM and potentially violate the assumption of the model.

4 Knock-out of BLK-FAM167A locus (rs558245864)

All primers and oligonulceotides are summarised in Table S1.

4.1 Construction of enhanced Cas9-2a-GFP vector

pSpCas9(BB)-2A-GFP (PX458) and eSpCas9(1.1) were gifts from Feng Zhang (Addgene plasmid # 48138 and 71814, respectively) and pMAX-GFP was purchased from Lonza, Cologne, Germany. To remove the hU6 promoter, gRNA scaffold and CAG enhancer sequences from PX459, the plasmid was digested with *PciI* and *AgeI* (New England Biolabs). The $\sim$8 kb fragment was agarose gel-purified. The primers pMAXcmvF and pMAXcmvR were used with 2x KAPA HiFi (Kapa Biosystems) to amplify a full length fragment of the CMV and chimeric intron, using pMAX-GFP as a template, followed also by agarose-gel purification. The PCR fragment containing the CMV and chimeric intron sequence was ligated into *PciI* and *AgeI*-digested PX459 using a Gibson Assembly reaction (New England Biolabs), according the manufacturer's instructions. Following propagation in NEB5 alpha competent bacteria (New England Biolabs), plasmid preparations were performed with the GenElute endotoxin-free midi kit (Sigma Aldrich). Plasmids were Sanger sequenced with the primers CMVmodQC and Cas9R1 to confirm correct insertion events. To convert the wildtype Cas9 sequence to the enhanced Cas9 sequence, the modified PX458 plasmid was digested with *EcoRV* and *BsmI* (New England Biolabs), releasing a fragment of the Cas9 coding sequence. The linearised plasmid was agarose-gel purified. To generate a replacement fragment spanning the three alanine substitutions in enhanced Cas9, the primers GIBeCas9FragF and GIBeCas9FragR were used with 2$\times$ KAPA HiFi to amplify the partial Cas9 sequence from eSpCas9(1.1). This fragment was ligated into the *EcoRV* and *BsmI* sites of the modified PX459 plasmid using a Gibson Assembly reaction, according to the manufacturer's instructions. Potential modified PX459 plasmids now carrying enhanced Cas9 were prepared as above, before Sanger sequencing was performed to confirm the correct modifications from wildtype Cas9 to enhanced Cas9.

4.2 Construction of rs558245864 targeting gRNA expression plasmid

The sgRNA expression plasmid p1261 was a kind gift from Sebastian Gerety, which was created based on the work by [26]. p1261 was digested with *BsaI* to remove the stuffer sequence that sits in between the human U6 promoter and the sgRNA backbone. *BsaI* digest of p1261 leaves a four-base 5′ overhang at either end for cloning in the target-specific sgRNA sequence. Two partially complementary target-bearing 24mer oligos, rs558245864topgRNA and rs558245864botgRNA, were ordered from Sigma Aldrich and were annealed before ligating into *BsaI*-digested p1261. Plasmids were prepared as above before Sanger sequence verification of correct cloning of target-specific sgRNA sequences relative to the human U6 promoter and sgRNA backbone.

4.3 Nucleofection and GFP sorting

The LCL cell line HG00146 was cultured as described previously [1]. Two days prior to nucleofection, cells at a density of $\sim 1.5 \times 10^6$ mL^{-1} were passaged with a one in three dilution. 2.5×10^6 cells were nucleofected with 6 μg of enhanced Cas9-2a-GFP, 0.83 μg of rs558245864 p1261 and 200 pmol of rs558245864 alternate allele single-stranded donor (Integrated DNA Technologies) using Nucleofector

Kit C (Lonza) with an Amaxa Nucelofector 2b (Lonza) on program X-001, according to the manufacturer's method. Cells were recovered in 5 mL of conditioned growth media. Conditioned media was made from an equal mix of fresh media and media collected from LCLs 48 hours after passage, followed by sterile filtration. Nucleofected cells were grown for 24 hours before sorting GFP-positive and propidium iodide-negative cells into conditioned media. For GFP sorting, the top 30 % of GFP positive cells were selected. This was about 1 % of the total cells that were nucleofected. GFP-positive cells were expanded for 7 days before clonal selection.

4.4 Clonal selection and expansion of clones

Single propidium iodide-negative cells were flow-sorted into round bottom tissue culture-treated 96 well cluster plates (Costar) containing 150 μL of conditioned media, prepared as for the nucleofection and supplemented with 0.25 μg mL^{-1} amphotericin B (Gibco). Plates were briefly spun and wells were examined under a microscope to identify wells with a single cell. Although very rare, wells with two or more cells were not expanded. Clonal cells were expanded with a 50 % media change with conditioned media every 5 days until enough cells were present to make duplicate 96-well V-bottom plates for genotyping.

4.5 Genotyping knock-out clones at the rs558245864 locus

Although an attempt was made to precisely engineer the alternate rs558245864 allele into HG00146, we were unsuccessful and only succeeded in generating knock-out clones of the rs558245864 locus. Knock-out clones were genotyped as follows: Duplicate 96-well V-bottom plates containing potential clones were spun briefly to pellet cells. The media was removed before 20 μL of a lysis buffer containing 25 mM NaOH and 0.2 mM Na_2-EDTA was added to the cells. Lysis was performed at 95 °C for 5 minutes to release the genomic DNA, before the addition of 20 μL of neutralisation buffer containing 40 mM Tris-HCL. Potential deletion clones were identified by performing a melt-curve analysis as follows: 2 μL of the genomic DNA was added to 12.5 μL of 2x KAPA SYBR Fast (Kapa Biosystems), 0.5 μL 10 μM rs558245864mcF primer, 0.5 μL 10 μM rs558245864mcR primer and 9.5 μL of nuclease-free water. Samples were amplified on an ABI StepOne Plus real-time PCR machine (Thermo Fisher) by incubating at 95 °C for 2 minutes, followed by 40 cycles of 95 °C for 20 seconds, 64 °C for 15 seconds and 72 °C for 15 seconds. Amplicons were then subject to a melt-curve dissociation ramping from 60 °C to 95 °C with data collection at 0.3 °C increments. Possible deletion clones were identified by comparing to amplicons generated from unedited cell lines and were then confirmed by Sanger sequencing as follows: 100 ng of genomic DNA from potential deletion clones was mixed with 12.5 μL of 2x KAPA HiFi PCR master mix, 1 μL 10 μM rs558245864seqF primer, 1 μL 10 μM rs558245864seqR primer and nuclease-free water, to a final volume of 25 μL. Samples were amplified by incubating at 95 °C for 5 minutes, followed by 27 cycles of 98 °C for 20 seconds, 65 °C for 20 seconds and 72 °C for 30 seconds. PCR products were agarose-gel purified before Sanger sequencing was performed (Eurofins Genomics) in both directions using rs558245864seqF and rs558245864seqR primers.

4.6 RNA-seq and ATAC-seq for rs558245864 knock-out and HG00142 lines

For each sample, total RNA was extracted from approximately 500,000 cells using the RNeasy Mini kit (Qiagen), including the on-column DNase digestion, according the manufacturer's instructions. RNA-seq libraries were prepared using the KAPA Stranded mRNA-seq kit for the Illumina platform (KAPABiosystems) with 500 ng of total RNA, according to the manufacturer's instructions. ATAC-seq on the same samples was performed as previously described [1]. Paired-end reads of 75 bp were generated on an Illumina HiSeq 4000 system for both ATAC-seq and RNA-seq. We obtained a total of 234 and 542 million fragments for ATAC-seq and RNA-seq on autosomes, respectively.

4.7 Differential chromatin accessibility and differential expression analyses

The ATAC-seq and RNA-seq data were processed in exactly the same way as described in Section 2 to obtain fragment counts for all the annotated peaks and annotated genes. For the ATAC-seq data, instead of calling peaks from the data itself, we used the peak annotation called from the 100 samples in the main study. We used DESeq [27] to perform differential chromatin accessibility and differential expression analyses. We compared the two replicates from the parental line and the four replicates from the knock-out lines. For the differential chromatin accessibility analysis, we used 214,584 peaks with mean FPKM greater than 0 and the total fragment counts greater than 100. For the differential expression analysis, we used 10,225 protein coding genes with mean FPKM greater than 0 and the total fragment counts greater than 100.

Table 1: rs558245864 engineering primer sequences

Oligo name	Sequence	Function
rs558245864 alternate allele single-stranded donor	*see footnote*[†]	Single-stranded donor molecule for precise engineering of alternate rs558245864 allele
pMAXcmvF	`cttttgctggccttttgctcatcaatattggccattagccat`	To amplify the full CMV and chimeric intron from pMAX
pMAXcmvR	`tccttatagtccatggtggcaccctgtggagagaaaggcaaag`	To amplify the full CMV and chimeric intron from pMAX
CMVmodQC	`cgccacctctgacttgagcgtcg`	Sequencing primer to confirm correct insertion of CMV promoter into PX459
Cas9R1	`gaacagcagggctccgatcag`	Sequencing primer to confirm correct modification of WT Cas9 to enhanced Cas9
Cas9F1	`cacgacggagactacaaggatc`	Sequencing primer to confirm correct modification of WT Cas9 to enhanced Cas9
Cas9F2	`cggctgatctatctggccctg`	Sequencing primer to confirm correct modification of WT Cas9 to enhanced Cas9
Cas9F3	`gcctctatgatcaagagatacg`	Sequencing primer to confirm correct modification of WT Cas9 to enhanced Cas9
Cas9F4	`ctgctgttcaagaccaaccgg`	Sequencing primer to confirm correct modification of WT Cas9 to enhanced Cas9
Cas9F5	`gacgagctcgtgaaagtgatg`	Sequencing primer to confirm correct modification of WT Cas9 to enhanced Cas9
Cas9F6	`gtgctgaccagaagcgacaag`	Sequencing primer to confirm correct modification of WT Cas9 to enhanced Cas9
Cas9F7	`gaccgagattaccctggccaa`	Sequencing primer to confirm correct modification of WT Cas9 to enhanced Cas9
Cas9F8	`cctgtacctggccagccacta`	Sequencing primer to confirm correct modification of WT Cas9 to enhanced Cas9
GIBeCas9FragF	`acgaggacattctggaagatatcgtgctgaccctgacactg`	To amplify a partial fragment of enhanced Cas9 carrying the three alanine substitutions
GIBeCas9FragR	`gcagttcgccggcagaggccagcattctcttccggccgtt`	To amplify a partial fragment of enhanced Cas9 carrying the three alanine substitutions
rs558245864topgRNA	`accgtgtgacagtctgccacacgg`	Top oligo for cloning rs558245864 sgRNA sequence into p1261
rs558245864botgRNA	`aaacccgtgtggcagactgtcaca`	Bottom oligo for cloning rs558245864 sgRNA sequence into p1261
U6F2	`tggactatcatatgcttaccg`	Sequencing primer for confirmation of correct sgRNA sequence into p1261
rs558245864mcF	`tttcttcttcccctggccct`	Forward primer for melt-curve analysis of potential rs558245842 locus deletion clones
rs558245864mcR	`gaagcccagttctcctgaggt`	Reverse primer for melt-curve analysis of potential rs558245842 locus deletion clones
rs558245864seqF	`tcctccctgctctcactccc`	Forward primer for Sanger sequencing confirmation of deletion clones
rs558245864seqR	`gtgaggggaggggtggaatg`	Reverse primer for Sanger sequencing confirmation of deletion clones

[†] `g*t*t*ataatgatggtctcaagtagtaattgtgacagtctgccacacgggggggggtcatgtcctacattacacagcatccccccaagg*g*g*a` (* indicates phosphorothioate linker)

Appendices

A. Student's t statistic to Z statistic conversion for Wakefield's approximation

The Wakefield's asymptotic Bayes factor approximation method assumes the input Z statistic follows the standard normal distribution. However, the Wald test statistic of the slope from the simple linear regression follows Student's t distribution, which violates the asymptotic normality especially when the sample size of the linear regression is small. In fact, the QQ-plot of approximated Bayes factors shows the tail of the distribution is strongly deviated from the analytic Bayes factors (blue dots in Fig. S10), suggesting the normal assumption of the Wakefield's method is not valid for the linear regression T statistic obtained from the 100 ATAC-seq samples.

Therefore we utilise the normal transformation $\mathcal{Z}(\cdot)$ (as in Appendix C of [28]) to convert the linear regression T statistic (from n samples) into the Z statistic, such as

$$Z = \mathcal{Z}(T) = \Phi^{-1}(F_{n-2}(T)),$$

where $\Phi(\cdot)$ and $F_n(\cdot)$ denote the cumulative density functions for the standard normal and Student's t distributions with n degrees of freedom, respectively. The transformation works perfectly for large Bayes factors obtained from the linear regression model (red dots in Fig. S10). Note that this transformation is easily implemented, for example, on `R`,

```
T # T statistic from linear regression model
n # sample size of linear regression model
Z = sign(T) * qnorm(pt(-abs(T), n-2)) # Z to T conversion
```

B. Two stage least square method (2SLS)

Mendelian Randomisation (MR) is a powerful technique to perform causal inference [22]. It is often used in Epidemiological studies to rigorously estimate the causal effect β_1 of exposure X on outcome Y, where various unknown confounding factors U hamper to perform the conventional randomised controlled trial (Fig. S11). Instrumental variable (IV) technique is one of the few ways to estimate causal effects without complete knowledge of all confounding factors of the exposure-outcome association. The instrumental variable G is a genetic variant that satisfies the following assumptions. Firstly, G has to be associated with X (IV1). Secondly, G is not associated with any confounding factor U of the exposure-outcome association (IV2). Finally, G can only affect Y through its association with X (IV3).

Under these assumptions, the two stage least square method (2SLS) [22] is often used to estimate β_1. It comprises two regression stages: the first-stage regression of X on G, and the second-stage regression of Y on the fitted values of X from the first stage. Suppose we have single IV with data on individuals $i = 1, \ldots, n$ who have exposure value x_i, outcome value y_i and genotype dosage value g_i. The first-stage regression model is

$$x_i = \alpha_0 1 + \alpha_1 g_i + \varepsilon_i \quad (i = 1, \ldots, n),$$

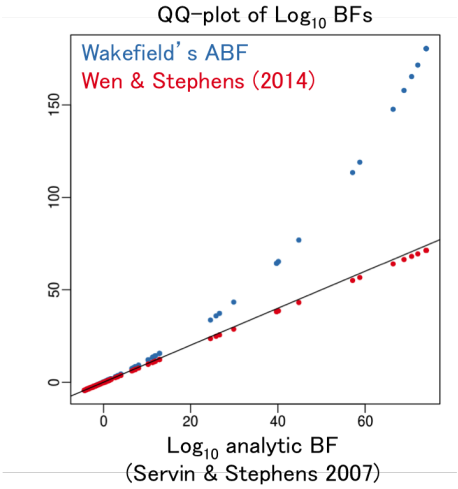


Fig. S10: QQ-plot shows the Wakefield's asymptotic Bayes factors (Y-axis) against the analytic Bayes factor (X-axis) obtained by [21]. The blue dots show the linear regression T statistics of caQTL mapping converted into Bayes factors by the Wakefield's method. The red dots show the Bayes factors calculated from the same T statistics, but obtained from the T to Z transformation [28] followed by the Wakefield's approximation. The data is based on the 100 ATAC-seq sample size. Here the variance parameter of the prior distribution for the slope was $W = 10$ for the Wakefield's approximation and $\sigma_a^2 = 10$ for the analytic Bayes factor.

where $\varepsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_X^2)$. Let us define the following sample means and (co)variances

$$\begin{aligned}
\mathrm{E}[G] &\equiv \frac{1}{n}\sum_{i=1}^{n} g_i, & \mathrm{Var}(G) &\equiv \frac{1}{n}\sum_{i=1}^{n}(g_i - \mathrm{E}[G])^2,\\
\mathrm{E}[X] &\equiv \frac{1}{n}\sum_{i=1}^{n} x_i, & \mathrm{Cov}(X,G) &\equiv \frac{1}{n}\sum_{i=1}^{n}(x_i - \mathrm{E}[X])(g_i - \mathrm{E}[G]),\\
\mathrm{E}[Y] &\equiv \frac{1}{n}\sum_{i=1}^{n} y_i, & \mathrm{Cov}(Y,G) &\equiv \frac{1}{n}\sum_{i=1}^{n}(y_i - \mathrm{E}[Y])(g_i - \mathrm{E}[G]),
\end{aligned}$$

we have the least square estimate

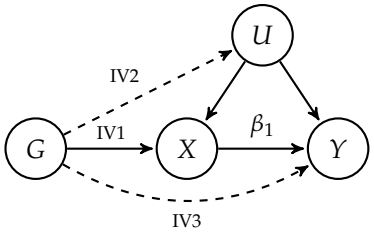
$$\begin{aligned}
\hat{\alpha}_0 &= \mathrm{E}[X] - \hat{\alpha}_1 \mathrm{E}[G],\\
\hat{\alpha}_1 &= \frac{\mathrm{Cov}(X,G)}{\mathrm{Var}(G)}.
\end{aligned}$$

Then the fitted values $\hat{x}_i = \hat{\alpha}_0 + \hat{\alpha}_1 g_i$ are used in the second stage regression model

$$y_i = \beta_0 1 + \beta_1 \hat{x}_i + \xi_i \quad (i = 1, \ldots, n),$$

where $\xi_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_Y^2)$ and it is also independent of ε_i. The least square estimate of the second-stage regression is given by

$$\begin{aligned}
\hat{\beta}_0 &= \mathrm{E}[Y] - \hat{\beta}_1 \mathrm{E}[X],\\
\hat{\beta}_1 &= \frac{\mathrm{Cov}(Y,G)}{\mathrm{Cov}(X,G)}.
\end{aligned}$$



G: instrumental variable
X: exposure
Y: outcome
U: confounding factors

Fig. S11: Schematic of Mendelian Randomisation setting. There are four key variables: instrumental variable (G), exposure (X), outcome (Y) and confounding factors (U). The arrow with solid line indicates two variables are associated and the arrow with broken line indicates two variables are marginally/conditionally independent. There are three assumption (IV1–IV3) that G has to satisfy.

Although the estimation of causal effect $\hat{\beta}_1$ gives the correct point estimate, the standard error from the second-stage regression is not correct. This is because it does not take into account the uncertainty in the first-stage regression. Under the homoscedasticity of the error term in the equation

$$y_i = \beta_0 1 + \beta_1 x_i + \xi_i \quad (i = 1, \ldots, n),$$

the variance estimator of ξ_i in the second-stage regression is given by

$$\hat{\sigma}_Y^2 = \frac{\sum_{i=1}^n \hat{\xi}_i^2}{n-2}$$

where $\hat{\xi}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$ ($\neq y_i - \hat{\beta}_0 - \hat{\beta}_1 \hat{x}_i$). Therefore the asymptotic variance of the 2SLS estimator is

$$\begin{aligned} \mathrm{Var}(\hat{\beta}_1) &= \frac{\hat{\sigma}_Y^2}{n\mathrm{Var}(\hat{X})} \\ &= \frac{\mathrm{Var}(Y - \hat{\beta}_1 X)}{(n-2)\mathrm{Var}(X)\mathrm{Cor}(X,G)^2}, \end{aligned}$$

where $\mathrm{Var}(\hat{X}) \equiv \sum_{i=1}^n (\hat{x}_i - \mathrm{E}[\hat{X}])^2/n$ and $\mathrm{E}[\hat{X}] \equiv \sum_{i=1}^n \hat{x}_i/n$.

C. Nonlinear model with cubic spline function

We use the spline smoothing technique to model nonlinearity of the prior probabilities in the generalised additive model framework [29]. The nonlinear function used in the hierarchical model

$$f(x;\beta) = \sum_{i=1}^{q} b_i(x)\beta_i$$

is described as a linear combination of q cubic spline basis functions $\{b_i(\cdot); i = 1, \ldots, q\}$ with some values of the unknown parameters $\{\beta_i; i = 1, \ldots, q\}$. Let the knot locations be denoted by $\{x_i^*; i = 1, \ldots, q-2\}$, a cubic spline basis is defined by $b_1(x) = 1$, $b_2(x) = x$, $b_{i+2}(x) = R(x, x_i^*)$ for $i = 1, \ldots, q-2$ where

$$R(x,z) = \frac{1}{4}\left[\left(z - \frac{1}{2}\right)^2 - \frac{1}{12}\right]\left[\left(x - \frac{1}{2}\right)^2 - \frac{1}{12}\right] - \frac{1}{24}\left[\left(|x-z| - \frac{1}{2}\right)^4 - \frac{1}{2}\left(|x-z| - \frac{1}{2}\right)^2 + \frac{7}{240}\right].$$

We normally introduce a penalty to secure the smoothness of f during the maximisation of likelihood, because it usually overfits to the data. The integrated square of second derivative

$$\int_0^1 f''(x)^2 dx = \beta^\top S \beta$$

is often used as a penalty on wiggliness of f. Because f is linear in the parameters $\beta = (\beta_1, \ldots, \beta_q)^\top$, the penalty can always be written as a quadratic form in β, where S is a matrix whose element is $S_{i+2,j+2} = R(x_i^*, x_j^*)$, for $i, j = 1, \ldots, q-2$ while the first two rows and columns of S are 0. (See Chapter 3.2.1 in [29] for more details).

We have chosen a set of equi-spaced knots in $[0, 1]$ region, such that

$$x_i^* = \frac{i - c}{(q-2) + 1 - 2c}$$

with the number of basis functions $q = 6$ and a constant $c = 3/8$ for the three levels of prior probability. Given the knot locations, for example, the linear predictor of the variant level prior probability is modelled by

$$\eta_{jl} = x_{\text{INDEL}}^{(l)} \lambda_1 + x_{\text{CNV}}^{(l)} \lambda_2 + \sum_{i=1}^{6} b_i(h_{al}) \lambda_{i+3},$$

when the variant l is inside a flanking peak $a \in \mathcal{W}_{jk}$ $(a \neq j)$.

D. Penalised iteratively reweighted least square on Θ_1

We begin with maximisation of the marginal likelihood in Eq. 6 with respect to $\Theta_1 = \{\Pi(\gamma), \pi(\lambda)\}$. Let Z_j be the indicator variable that the peak j is a QTL and $z_j^{(l)}$ be the indicator variable that the variant $l \in \mathcal{W}_j$ is the putative causal caQTL for the peak j such that $\sum_{l \in \mathcal{W}_j} z_j^{(l)} = 1$. The complete log likelihood of the marginal structure is written by

$$l_1^c(\Theta_1) = \sum_{j=1}^{J} \left[(1 - Z_j) \log \Pi_j^{(0)} + Z_j \log \Pi_j^{(1)} + Z_j \sum_{l \in \mathcal{W}_j} z_j^{(l)} \log \pi_j^{(l)} BF_j^{(l)} \right] - \frac{1}{2} \lambda^\top S_1 \lambda - \frac{1}{2} \gamma^\top S_2 \gamma, \quad (13)$$

where the penalties of wiggliness for f_1 and f_2 are given by quadratic forms with $\lambda = (\lambda_1, \ldots, \lambda_9)^\top$ and $\gamma = (\gamma_1, \ldots, \gamma_6)^\top$, respectively (see Appendix C for detail.). The penalty matrices are defined by

$$S_1 = \nu_1 \begin{pmatrix} 0 & 0 \\ 0 & S \end{pmatrix} + \tau I_9 \quad \text{and} \quad S_2 = \nu_2 S + \tau I_6$$

where S is given in Appendix C and $\{\nu_1 = 5.0, \nu_2 = 1.0\}$ are the smoothing parameters. Here we also introduce a ridge-regression type penalty $\tau = 0.01$ to address an ill-conditioned problem, such as multiple optimum parameters (I_n is the n-dimensional identity matrix).

We use the Expectation-Maximisation (EM) algorithm [24] to estimate the parameters. In E-step, the complete likelihood in Eq. 13 is averaged over the posterior distribution of missing variables Z_j and $z_j = (z_j^{(l)}; l \in \mathcal{W}_j)^\top$ under the current estimate of parameters $\hat{\Theta}_1$ to obtain the Q function, such that $Q(\Theta_1 | \hat{\Theta}_1) = \mathbb{E}[l_1^c(\Theta_1) | \prod_j p(Z_j, z_j | Y, \hat{\Theta}_1)]$. We can easily obtain the Q function from the complete likelihood Eq. 13 by simply replacing Z_j and z_j with

$$\overline{Z}_j = \frac{\hat{\Pi}_j^{(1)} R\hat{B}F_j}{\hat{\Pi}_j^{(0)} + \hat{\Pi}_j^{(1)} R\hat{B}F_j} \quad \text{and} \quad \overline{z}_j^{(l)} = \frac{\hat{\pi}_j^{(l)} BF_j^{(l)}}{R\hat{B}F_j} \quad (14)$$

for $j = 1, \ldots, J$, where $R\hat{B}F_j = \sum_{m \in \mathcal{W}_j} \hat{\pi}_j^{(m)} BF_j^{(m)}$.

In the M-step, the maximisation of Q function is performed alternately with λ and γ. Let

$$x_{jl} = \begin{cases} (x_{\text{INDEL}}^{(l)}, x_{\text{CNV}}^{(l)}, 1, 0, b_2(h_{jl}), \ldots, b_6(h_{jl}))^\top & l \text{ inside the focal peak } j \\ (x_{\text{INDEL}}^{(l)}, x_{\text{CNV}}^{(l)}, 0, 1, b_2(h_{al}), \ldots, b_6(h_{al}))^\top & l \text{ inside a flanking peak } a \in \mathcal{W}_j (\neq j) \\ (x_{\text{INDEL}}^{(l)}, x_{\text{CNV}}^{(l)}, 0, 0, 0, \ldots, 0)^\top & \text{otherwise} \end{cases} \tag{15}$$

be the covariates of the variant level prior probability with the relative coverage h_{jl} or h_{al} at variant l, so that

$$\pi_j^{(l)} = \frac{\exp(x_{jl}^\top \lambda)}{\sum_{m \in \mathcal{W}_j} \exp(x_{jm}^\top \lambda)}.$$

The first and second derivative of Q with respect to λ are given by

$$\left.\frac{\partial Q}{\partial \lambda^\top}\right|_{\lambda=\hat{\lambda}} = \sum_{j=1}^{J} \overline{Z}_j \sum_{l \in \mathcal{W}_j} \bar{z}_j^{(l)} \tilde{x}_{jl} - S_1 \hat{\lambda} = X_1^\top \bar{z} - S_1 \hat{\lambda}$$

$$\left.\frac{\partial^2 Q}{\partial \lambda \partial \lambda^\top}\right|_{\lambda=\hat{\lambda}} = -\sum_{j=1}^{J} \overline{Z}_j \sum_{l \in \mathcal{W}_j} \hat{\pi}_j^{(l)} \tilde{x}_{jl} \tilde{x}_{jl}^\top - S_1 = -X_1^\top W_1 X_1 - S_1$$

where $\tilde{x}_{jl} = x_{jl} - \sum_{l \in \mathcal{W}_j} \hat{\pi}_j^{(l)} x_{jl}$,

$$\bar{z} = \begin{pmatrix} \overline{Z}_1 \bar{z}_1 \\ \vdots \\ \overline{Z}_J \bar{z}_J \end{pmatrix}, \; X_1 = \begin{pmatrix} (\tilde{x}_{1l}; l \in \mathcal{W}_1)^\top \\ \vdots \\ (\tilde{x}_{Jl}; l \in \mathcal{W}_J)^\top \end{pmatrix} \text{ and } \; W_1 = \operatorname{diag} \begin{pmatrix} \overline{Z}_1 \hat{\pi}_1 \\ \vdots \\ \overline{Z}_J \hat{\pi}_J \end{pmatrix}.$$

Note that $(\tilde{x}_{jl}; l \in \mathcal{W}_j)$ is a matrix whose column is $\tilde{x}_{jl}$ for all variants l in $\mathcal{W}_j$ and $\hat{\pi}_j = (\hat{\pi}_j^{(l)}; l \in \mathcal{W}_j)^\top$. The result implies that maximisation of the Q function does not require to compute the actual gradient and hessian for an iterative optimisation approach, such as the Newton-Raphson method. Instead, each Newton-Raphson step is equivalent to minimise the following weighted squared norm

$$\|y_1^* - X_1 \lambda\|_{W_1}^2 + \lambda^\top S_1 \lambda$$

with respect to λ, where $y_1^* = X_1 \lambda + W_1^{-1} \bar{z}$ is so called the pseudodata as if it is a response variable of the standard linear regression (see Chapter 3.4 of [29] for details). The solution to attain the minimum is obtained by

$$\hat{\lambda}_{\text{new}} = R_1^{-1} Q_1^\top W_1^{\frac{1}{2}} y_1^*, \tag{16}$$

where the QR decomposition of the following expanded matrix

$$\begin{pmatrix} W_1^{\frac{1}{2}} X_1 \\ C_1 \end{pmatrix} = Q_1 R_1$$

is used. Here C_1 is a square root matrix such that $S_1 = C_1^\top C_1$ (*e.g.*, obtained by Cholesky decomposition). This optimisation approach is referred to as the penalised iteratively reweighted least square (P-IRLS) method.

In analogy with the optimisation with respect to λ, P-IRLS is also used to maximise the Q function with respect to γ. Let

$$u_j = (b_1(q_j), \ldots, b_6(q_j))^\top$$

be the covariates for the peak level prior probability for peak j with the peak height quantile q_j, so that

$$\Pi_j^{(1)} = \frac{\exp(u_j^\top \gamma)}{1 + \exp(u_j^\top \gamma)}.$$

The first and second derivatives of Q with respect to γ are obtained by

$$\left.\frac{\partial Q}{\partial \gamma^\top}\right|_{\gamma=\hat{\gamma}} = \sum_{j=1}^{J} (\overline{Z}_j - \hat{\Pi}_j^{(1)}) u_j - S_2\hat{\gamma} = X_2^\top(\overline{Z} - \hat{\Pi}^{(1)}) - S_2\hat{\gamma}$$

$$\left.\frac{\partial^2 Q}{\partial \gamma \partial \gamma^\top}\right|_{\gamma=\hat{\gamma}} = -\sum_{j=1}^{J} \hat{\Pi}_j^{(0)}\hat{\Pi}_j^{(1)} u_j u_j^\top - S_2 = -X_2{}^\top W_2 X_2 - S_2,$$

where

$$\overline{Z} = \begin{pmatrix} \overline{Z}_1 \\ \vdots \\ \overline{Z}_J \end{pmatrix}, \ \hat{\Pi}^{(1)} = \begin{pmatrix} \hat{\Pi}_1^{(1)} \\ \vdots \\ \hat{\Pi}_J^{(1)} \end{pmatrix}, \ X_2 = \begin{pmatrix} u_1^\top \\ \vdots \\ u_J^\top \end{pmatrix} \text{ and } W_2 = \text{diag} \begin{pmatrix} \hat{\Pi}_1^{(0)}\hat{\Pi}_1^{(1)} \\ \vdots \\ \hat{\Pi}_J^{(0)}\hat{\Pi}_J^{(1)} \end{pmatrix}.$$

Again, each Newton-Raphson step is equivalent to minimise the weighted squared norm

$$\|y_2^* - X_2\gamma\|_{W_2}^2 + \gamma^\top S_2 \gamma$$

with respect to γ, where $y_2^* = X_2\gamma + W_2^{-1}(\overline{Z} - \hat{\Pi}^{(1)})$ is the pseudodata. The solution to attain the minimum is obtained by

$$\hat{\gamma}_{\text{new}} = R_2^{-1} Q_2^\top W_2^{\frac{1}{2}} y_2^*,$$

where the QR decomposition of the expanded matrix

$$\begin{pmatrix} W_2^{\frac{1}{2}} X_2 \\ C_2 \end{pmatrix} = Q_2 R_2$$

is used. Here C_2 is a square root matrix such that $S_2 = C_2^\top C_2$. This alternate P-IRLS step for λ and γ is performed once every M-step until arriving at a convergence.

In order to compute the standard error for $\hat{\lambda}$ and $\hat{\gamma}$, we use the empirical observed information matrix (see Chapter 4.3 of [24] for details). Let us denote the expected score vector with respect to λ and γ for peak j by

$$s_j^\lambda = \left.\frac{\partial Q_j}{\partial \lambda^\top}\right|_{\lambda=\hat{\lambda}} = \overline{Z}_j \sum_{l \in \mathcal{W}_j} \overline{z}_j^{(l)} \tilde{x}_{jl},$$

$$s_j^\gamma = \left.\frac{\partial Q_j}{\partial \gamma^\top}\right|_{\gamma=\hat{\gamma}} = (\overline{Z}_j - \hat{\Pi}_j^{(1)}) u_j,$$

where Q_j is a part of the Q function corresponding to peak j so that $Q = \sum_{j=1}^{J} Q_j$. Then we have

$$\text{Var}(\hat{\lambda}) \approx \left[\sum_{j=1}^{J} s_j^\lambda (s_j^\lambda)^\top\right]^{-1} \quad \text{and} \quad \text{Var}(\hat{\gamma}) \approx \left[\sum_{j=1}^{J} s_j^\gamma (s_j^\gamma)^\top\right]^{-1}.$$

E. Penalised iteratively reweighted least square on Θ_2

By using the estimate of parameters $\hat{\Theta}_1 = \{\Pi(\hat{\gamma}), \pi(\hat{\lambda})\}$ we then maximise the pairwise likelihood in Eq. 7 with respect to $\Theta_2 = \{\Psi(\delta)\}$. According to the full set of interaction hypotheses demonstrated in Eq. 5, let $Z_{jk}^{(h)}$ be the indicator variable which is $Z_{jk}^{(h)} = 1$ if interaction hypothesis $h \in \mathcal{H}$ is true for the j–k peak pair; otherwise $Z_{jk}^{(h)} = 0$. Then the complete log likelihood of the pairwise structure is written as

$$l_2^c(\Theta_2; \hat{\Theta}_1) = \sum_{\substack{1 \le j < k \le J \\ d(j,k) < 5 \times 10^5}} \left[\sum_{h \in \mathcal{H}_0} Z_{jk}^{(h)} \log \Phi_{jk}^{(h)} + \sum_{h \in \mathcal{H}_1} Z_{jk}^{(h)} \log \Phi_{jk}^{(h)} R\hat{B}F_{jk}^{(h)} \right] - \frac{1}{2} \delta^\top S_3 \delta, \tag{17}$$

where the estimate of regional Bayes factors $\{R\hat{B}F_{jk}^{(h)}\}$ is followed from Eq. 4 with the estimate of variant level prior probabilities $\pi(\hat{\lambda})$. The penalty matrix is defined by

$$S_3 = \nu_3 \begin{pmatrix} S & 0 \\ 0 & S \end{pmatrix} + \tau I_{12}$$

with S defined in Appendix C and the ridge regression penalty $\tau = 0.01$.

We use the EM-algorithm to maximise Eq. 17. In the E-step, we compute the posterior probability of each interaction hypothesis

$$\overline{Z}_{jk}^{(h)} = \begin{cases} \dfrac{\hat{\Phi}_{jk}^{(h)}}{\sum_{i \in \mathcal{H}_0} \hat{\Phi}_{jk}^{(i)} + \sum_{i \in \mathcal{H}_1} \hat{\Phi}_{jk}^{(i)} R\hat{B}F_{jk}^{(i)}} & h \in \mathcal{H}_0 \\ \dfrac{\hat{\Phi}_{jk}^{(h)} R\hat{B}F_{jk}^{(h)}}{\sum_{i \in \mathcal{H}_0} \hat{\Phi}_{jk}^{(i)} + \sum_{i \in \mathcal{H}_1} \hat{\Phi}_{jk}^{(i)} R\hat{B}F_{jk}^{(i)}} & h \in \mathcal{H}_1 \end{cases} \tag{18}$$

under the current estimate of parameters $\hat{\Theta}_2$. The posterior probability is used to compute Q function regarding the peak-pair level prior probabilities $\{\Psi_{jk}^{(i)}; i = 0, \ldots, 3\}$ in the M-step, such that

$$Q(\Theta_2 | \hat{\Theta}_2) = \sum_{\substack{1 \le j < k \le J \\ d(j,k) < 5 \times 10^5}} \left[\sum_{i=0}^{3} \overline{U}_{jk}^{(i)} \log \Psi_{jk}^{(i)} \right] - \frac{1}{2} \delta^\top S_3 \delta + \text{const}, \tag{19}$$

where

$$\overline{U}_{jk}^{(i)} = \begin{cases} \overline{Z}_{jk}^{(0.0)} + \overline{Z}_{jk}^{(1.1)} + \overline{Z}_{jk}^{(1.2)} + \overline{Z}_{jk}^{(2)} & i = 0 \text{ (no interaction)} \\ \overline{Z}_{jk}^{(0.1)} + \overline{Z}_{jk}^{(3)} & i = 1 \text{ (pleiotropy)} \\ \overline{Z}_{jk}^{(0.2)} + \overline{Z}_{jk}^{(4.1)} & i = 2 \text{ (causality } j \to k) \\ \overline{Z}_{jk}^{(0.3)} + \overline{Z}_{jk}^{(4.2)} & i = 3 \text{ (causality } k \to j). \end{cases}$$

Let

$$V_{jk}^\top = \begin{pmatrix} v_{jk}^{(1)} \\ v_{jk}^{(2)} \\ v_{jk}^{(3)} \end{pmatrix} = \begin{pmatrix} b_1(d_{jk}) & \ldots & b_6(d_{jk}) & 0 & \ldots & 0 \\ 0 & \ldots & 0 & b_1(d_{jk}) & \ldots & b_6(d_{jk}) \\ 0 & \ldots & 0 & b_1(d_{jk}) & \ldots & b_6(d_{jk}) \end{pmatrix}$$

be the covariates for the peak level prior probability for peak j with the peak pair distance d_{jk}, so that

$$\Psi_{jk}^{(0)} = \frac{1}{1+\exp(v_{jk}^{(1)}\delta)+\exp(v_{jk}^{(2)}\delta)+\exp(v_{jk}^{(3)}\delta)},$$
$$\Psi_{jk}^{(i)} = \frac{\exp(v_{jk}^{(i)}\delta)}{1+\exp(v_{jk}^{(1)}\delta)+\exp(v_{jk}^{(2)}\delta)+\exp(v_{jk}^{(3)}\delta)},$$

for $i = 1, 2, 3$. Note that $v_{jk}^{(i)}$ $(i = 0, \ldots, 3)$ are row vectors. Using the following vector notations

$$\overline{U}_{jk}^{(-0)} = \begin{pmatrix} \overline{U}_{jk}^{(1)} \\ \overline{U}_{jk}^{(2)} \\ \overline{U}_{jk}^{(3)} \end{pmatrix} \text{ and } \hat{\Psi}_{jk}^{(-0)} = \begin{pmatrix} \hat{\Psi}_{jk}^{(1)} \\ \hat{\Psi}_{jk}^{(2)} \\ \hat{\Psi}_{jk}^{(3)} \end{pmatrix},$$

the first and second derivatives of Q with respect to δ are given by

$$\begin{aligned}
\left.\frac{\partial Q}{\partial \delta^\top}\right|_{\delta=\hat{\delta}} &= \sum_{\substack{1\le j<k\le J \\ d(j,k)<5\times 10^5}} V_{jk}(\overline{U}_{jk}^{(-0)} - \hat{\Psi}_{jk}^{(-0)}) - S_3\hat{\delta} \\
&= X_3^\top(\overline{U}^{(-0)} - \hat{\Psi}^{(-0)}) - S_3\hat{\delta} \\
\left.\frac{\partial^2 Q}{\partial\delta\partial\delta^\top}\right|_{\delta=\hat{\delta}} &= -\sum_{\substack{1\le j<k\le J \\ d(j,k)<5\times 10^5}} V_{jk}\left[\operatorname{diag}\left(\hat{\Psi}_{jk}^{(-0)}\right) - \hat{\Psi}_{jk}^{(-0)}\left(\hat{\Psi}_{jk}^{(-0)}\right)^\top\right] V_{jk}^\top - S_3 \\
&= -\sum_{\substack{1\le j<k\le J \\ d(j,k)<5\times 10^5}} V_{jk}P_{jk}V_{jk}^\top - S_3 \\
&= -X_3^\top W_3 X_3 - S_3
\end{aligned}$$

where

$$\overline{U}^{(-0)} = \begin{pmatrix} \overline{U}_{1,2}^{(-0)} \\ \vdots \\ \overline{U}_{J-1,J}^{(-0)} \end{pmatrix}, \ \hat{\Psi}^{(-0)} = \begin{pmatrix} \hat{\Psi}_{1,2}^{(-0)} \\ \vdots \\ \hat{\Psi}_{J-1,J}^{(-0)} \end{pmatrix}, \ X_3 = \begin{pmatrix} V_{1,2}^\top \\ \vdots \\ V_{J-1,J}^\top \end{pmatrix} \text{ and } W_3 = \begin{pmatrix} P_{1,2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & P_{J-1,J} \end{pmatrix}.$$

Note that the vectors and matrices above repeat the same structure across all j–k peak pairs, such that $1 \le j < k \le J$ and $d(j,k) < 500{,}000$. W_3 is a block diagonal matrix. This fact suggests, each Newton-Raphson step is equivalent to minimise the weighted squared norm

$$\|y_3^* - X_3\delta\|_{W_3}^2 + \delta^\top S_3\delta$$

with respect to δ, where $y_3^* = X_3\delta + W_3^{-1}(\overline{U}^{(-0)} - \hat{\Psi}^{(-0)})$ is the pseudodata. The solution to attain the minimum is obtained by

$$\hat{\delta}_{\text{new}} = R_3^{-1}Q_3^\top W_3^{\frac{1}{2}} y_3^*,$$

where the QR decomposition of the expanded matrix

$$\begin{pmatrix} W_3^{\frac{1}{2}}X_3 \\ C_3 \end{pmatrix} = Q_3R_3$$

is used. Here C_3 is a square root matrix such that $S_3 = C_3^\top C_3$. In practice, it is not suitable to perform Cholesky Decomposition on W_3 at every M-step. There is a analytic form for each block diagonal element

$$P_{jk}^{\frac{1}{2}} = \begin{pmatrix} \sqrt{(1-\hat{\Psi}_{jk}^{(1)})\hat{\Psi}_{jk}^{(1)}} & -\frac{\hat{\Psi}_{jk}^{(2)}\sqrt{\hat{\Psi}_{jk}^{(1)}}}{\sqrt{1-\hat{\Psi}_{jk}^{(1)}}} & -\frac{\hat{\Psi}_{jk}^{(3)}\sqrt{\hat{\Psi}_{jk}^{(1)}}}{\sqrt{1-\hat{\Psi}_{jk}^{(1)}}} \\ 0 & \sqrt{\hat{\Psi}_{jk}^{(2)} - \frac{(\hat{\Psi}_{jk}^{(2)})^2}{1-\hat{\Psi}_{jk}^{(1)}}} & -\frac{\sqrt{\hat{\Psi}_{jk}^{(2)}}\hat{\Psi}_{jk}^{(3)}}{\sqrt{(1-\hat{\Psi}_{jk}^{(1)})(1-\hat{\Psi}_{jk}^{(1)}-\hat{\Psi}_{jk}^{(2)})}} \\ 0 & 0 & \sqrt{\hat{\Psi}_{jk}^{(3)} - \frac{(\hat{\Psi}_{jk}^{(3)})^2}{1-\hat{\Psi}_{jk}^{(1)}-\hat{\Psi}_{jk}^{(2)}}} \end{pmatrix}.$$

This square root matrix is also useful to compute the pseudodata with the back-solve algorithm.

In order to compute the standard error for $\hat{\delta}$, we use the empirical observed information matrix (see Chapter 4.3 of [24] for details). Let us denote the expected score vector with respect to δ for j–k peak pair by

$$s_{jk}^{\delta} = \frac{\partial Q_{jk}}{\partial \delta^\top}\bigg|_{\delta=\hat{\delta}} = V_{jk}(\overline{U}_{jk}^{(-0)} - \hat{\Psi}_{jk}^{(-0)}),$$

where Q_{jk} is a part of the Q function corresponding to j–k peak pair so that $Q = \sum_{j,k} Q_{jk}$. Then we have

$$\mathrm{Var}(\hat{\delta}) \approx \left[\sum_{\substack{1\le j<k\le J \\ d(j,k)<5\times 10^5}} s_{jk}^{\delta}(s_{jk}^{\delta})^\top \right]^{-1}.$$

Note that, the estimated standard error could be underestimated because: (1) the uncertainty in the marginal structure (*i.e.*, standard error of $\hat{\lambda}$ and $\hat{\gamma}$) is not taken into account; and (2) in reality, j–k peak pair is not independent from j–k' peak pair because summary statistics of peak j are shared.

F. Penalised iteratively reweighted least square for eQTL mapping

We maximise the marginal likelihood in Eq. 10 with respect to $\Theta = \{\Pi(\gamma), \pi(\lambda)\}$. The complete likelihood is essentially the same as in Eq. 13 where Z_k is the indicator variable that the gene k is an eQTL and $z_k^{(l)}$ is the indicator variable that the variant $l \in \mathcal{W}_k$ is the putative causal eQTL for the gene k. For the annotation comparison, the covariates for the variant level prior defined in Eq. 15 is written as

$$x_{kl} \equiv \begin{cases} b_{kl} & \text{HiChIP} \\ a_l b_{kl} & \text{HiChIP+ATAC} \\ c_{kl} & \text{CHi-C} \\ a_l & \text{ATAC} \\ d_l & \text{ATAC+VL} \\ a_l c_{kl} & \text{ATAC+CHi-C} \\ (\log w_l, a_l)^\top & \text{ATAC+PMR} \\ (\log w_l, d_l^\top)^\top & \text{ATAC+PMR+VL} \end{cases}$$

with corresponding coefficients

$$\lambda \equiv \begin{cases} (\lambda_1, \ldots, \lambda_6)^\top & \text{ATAC+VL} \\ (\phi, \lambda)^\top & \text{ATAC+PMR} \\ (\phi, \lambda_1, \ldots, \lambda_6)^\top & \text{ATAC+PMR+VL} \end{cases}$$

Here

$$d_l = \begin{cases} (b_1(h_{jl}), \ldots, b_6(h_{jl}))^\top & l \text{ inside the peak } j \in \mathcal{W}_k \\ (0, \ldots, 0)^\top & \text{otherwise} \end{cases}$$

and $b_i(\cdot)$, $i = 1, \ldots, 6$ are the B-spline bases defined in Appendix C.

In the M-step, the maximisation of Q function with respect to λ is identical to minimising

$$\|y_1^* - X_1\lambda\|_{W_1}^2 + \tau I$$

with respect to λ where y_1^* denotes the pseudodata, W_1 denotes the diagonal matrix of weights and τI denotes the ridge type penalty, defined in Appendix D. The solution to attain the minimum is obtained by Eq. 16 with the QR decomposition of the expanded model matrix

$$\begin{pmatrix} W_1^{\frac{1}{2}} X_1 \\ \sqrt{\tau} I \end{pmatrix} = Q_1 R_1.$$

Note that the E-step and the M-step with respect to γ are exactly same as in Appendix D.

For the full annotation model, we define the covariates

$$x_{kl} = (\log w_{kl}, x_{\text{INDEL}}^{(l)}, x_{\text{CNV}}^{(l)}, d_l^\top)^\top$$

with $\lambda \equiv (\phi, \lambda_1, \ldots, \lambda_8)^\top$. Then the identical EM-algorithm in Appendix D is applicable.

G. EM algorithm for colocalisation model

We split the mixture probability for the no QTL pair (H_0) into two sub-hypotheses: (0) non-pleiotropic; or (1) pleiotropic, such that

$$\Phi_{jk}^{(h)} = \begin{cases} \Psi^{(0)}\Pi^{(0)}\Xi^{(0)} & h = H_{0.0} \\ \Psi^{(1)}\Delta^{(0)} & h = H_{0.1}, \end{cases}$$

then the E-step is identical to Eq. 18 with $\mathcal{H}_0 = \{H_{0.0}, H_{0.1}\}$ and $\mathcal{H}_1 = \{H_{1.1}, H_{1.2}, H_2, H_3\}$.

The M-step does not require any iterative maximisation approach and the parameter estimate is obtained only from the posterior probability $\overline{Z}_{jk}^{(h)}$. Each M-step updates the four parameters as follows:

$$\hat{\Psi}^{(1)} = \frac{\sum_{j,k}(\overline{Z}_{jk}^{(0.1)} + \overline{Z}_{jk}^{(3)})}{\sum_{j,k}\sum_{h\in\mathcal{H}}\overline{Z}_{jk}^{(h)}},$$

$$\hat{\Pi}^{(1)} = \frac{\sum_{j,k}(\overline{Z}_{jk}^{(1.1)} + \overline{Z}_{jk}^{(2)})}{\sum_{j,k}(\overline{Z}_{jk}^{(0.0)} + \overline{Z}_{jk}^{(1.1)} + \overline{Z}_{jk}^{(1.2)} + \overline{Z}_{jk}^{(2)})},$$

$$\hat{\Xi}^{(1)} = \frac{\sum_{j,k}(\overline{Z}_{jk}^{(1.2)} + \overline{Z}_{jk}^{(2)})}{\sum_{j,k}(\overline{Z}_{jk}^{(0.0)} + \overline{Z}_{jk}^{(1.1)} + \overline{Z}_{jk}^{(1.2)} + \overline{Z}_{jk}^{(2)})},$$

$$\hat{\Delta}^{(1)} = \frac{\sum_{j,k}\overline{Z}_{jk}^{(3)}}{\sum_{j,k}(\overline{Z}_{jk}^{(0.1)} + \overline{Z}_{jk}^{(3)})}.$$

References

[1] Kumasaka N, Knights AJ, Gaffney DJ (2016) Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. Nat Genet 48: 206-13.

[2] Jiang H, Lei R, Ding SW, Zhu S (2014) Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics 15.

[3] Li H, Durbin R (2009) Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics 25: 1754-60.

[4] Lappalainen T, Sammeth M, Friedländer MR, 't Hoen PA, Monlong J, et al. (2013) Transcriptome and genome sequencing uncovers functional variation in humans. Nature 501: 506-11.

[5] Ding Z, Ni Y, Timmer SW, Lee BK, Battenhouse A, et al. (2014) Quantitative genetics of CTCF binding reveal local sequence effects and different modes of X-chromosome association. PLoS Genet : in press.

[6] Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9: 357-9.

[7] Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, et al. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14: R36.

[8] Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, et al. (2010) Understanding mechanisms underlying human gene expression variation with rna sequencing. Nature 464: 768-72.

[9] Alter O, Brown PO, Botstein D (2000) Singular value decomposition for genome-wide expression data processing and modeling. Proc Natl Acad Sci USA 97: 10101–10106.

[10] Browning BL, Browning SR (2013) Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics 194: 459-71.

[11] Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, et al. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159: 1665–1680.

[12] Mumbach MR, Satpathy AT, Boyle EA, Dai C, Gowen BG, et al. (2017) Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. Nat Genet .

[13] Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, et al. (2016) Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst 3: 95–98.

[14] Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, et al. (2015) Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet 47: 598–606.

[15] Cairns J, Freire-Pritchett P, Wingett SW, Varnai C, Dimond A, et al. (2016) CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. Genome Biol 17: 127.

[16] Hoffman MM, Buske OJ, Wang J, Weng Z, Bilmes JA, et al. (2012) Unsupervised pattern discovery in human chromatin structure through genomic segmentation. Nat Methods 9: 473–476.

[17] Ernst J, Kellis M (2012) ChromHMM: automating chromatin-state discovery and characterization. Nat Methods 9: 215–216.

[18] Wakefield J (2010) Bayesian methods for examining hardy-weinberg equilibrium. Biometrics 66: 257-65.

[19] Pickrell JK (2014) Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. Ann Appl Stat 94: 559–73.

[20] Cox DR, Reid N (2004) Partial likelihood. Biometrika 62: 269–276.

[21] Servin B, Stephens M (2007) Imputation-based analysis of association studies: candidate regions and quantitative traits. PLoS Genet 3: e114.

[22] Burgess S, Thompson SG (2015) Mendelian Randomization. Boca Raton: Chapman & Hall/CRC.

[23] Veyrieras JB, Kudaravalli S, Kim SY, Dermitzakis ET, Gilad Y, et al. (2008) High-resolution mapping of expression-qtls yields insight into human gene regulation. PLoS Genet 4: e1000214.

[24] McLachlan GJ, Krishman T (1997) The EM Algorithm and Extensions. New York: John Wiley & Sons.

[25] Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, et al. (2014) Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 10: e1004383.

[26] Mali P, Yang L, Esvelt KM, Aach J, Guell M, et al. (2013) RNA-guided human genome engineering via Cas9. Science 339: 823–6.

[27] Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11: R106.

[28] Wen X, Stephens M (2014) Bayesian methods for genetic association analysis with heterogeneous subgroups: From meta-analyses to gene-environment interactions. Ann Appl Stat 8: 176–203.

[29] Wood SN (2006) Generalized Additive Models. Boca Raton: Chapman & Hall/CRC.