Shared and distinct molecular effects of regulatory genetic variants provide insight into mechanisms of distal enhancer-promoter communication

Helen Ray-Jones[1,2,3,4,*], Zeji Song[1,2,§], Alexander Haglund[5], Pavel Artemov[1,2], Monica Della Rosa[1,2,§], Frances Burden[6,7,§], Roman Kreuzhuber[6,7,8,§], Anna Litovskikh[1,2,§], Vanessa Xue Hui Tan[1,2,§], Lai Ting Chan[3,4], Mattia Frontini[6,7,9], Chris Wallace[10,11], Valeriya Malysheva[1,2,3,4,#], Leonardo Bottolo[11,12,13,†,#], Elena Vigorito[11,†,#] and Mikhail Spivakov[1,2,#,*,^]

^Lead contact: M.S.
*Corresponding authors: H.R.J. and M.S.
†Equal contribution, listed alphabetically: L.B. and E.V.
#Senior authors: V.M., L.B., E.V and M.S.

1 - MRC London Institute of Medical Sciences, London W12 0NN, UK
2 - Institute of Clinical Sciences, Imperial College Faculty of Medicine, London W12 0NN
3 - Computational Neurobiology, VIB Center for Molecular Neurology, VIB, Antwerp, Belgium
4 - Computational Neurobiology, Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium
5 - Department of Brain Sciences, Faculty of Medicine, Imperial College London, London, UK
6 - Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK
7 - National Health Service (NHS) Blood and Transplant, Cambridge Biomedical Campus, Cambridge, UK
8 - EMBL-EBI, Wellcome Genome Campus, Cambridge, UK
9 - Department of Clinical and Biomedical Sciences, Faculty of Health and Life Sciences, University of Exeter Medical School, Exeter, UK
10 - Cambridge Institute of Therapeutic Immunology & Infectious Disease (CITIID), Jeffrey Cheah Biomedical Centre, University of Cambridge, Cambridge, CB2 0AW, UK
11 - MRC Biostatistics Unit, School of Clinical Medicine, University of Cambridge, Cambridge, UK
12 - Department of Medical Genetics, School of Clinical Medicine, University of Cambridge, Cambridge, UK
13 - The Alan Turing Institute, London, UK

§Present addresses: MDR: Cyted, Cambridge, UK; FB: University of Kent, Canterbury, UK; RK: Swiss Federal Administration, Bern, Switzerland; AL: Institute of Computational Biology, Helmholtz Zentrum München and Ludwig Maximilians University Munich, Faculty of Medicine, Munich, Germany; VXHT: Hummingbird Bioscience, Singapore.

Abstract

Transcriptional enhancers regulate gene expression in time and space, commonly engaging in long-range chromosomal contacts with gene promoters. However, the relationship between enhancer activity, enhancer-promoter contacts and gene expression is not fully understood. Here, we leveraged human genetic variation as a "natural perturbation" to dissect this relationship, focusing on distal enhancers containing expression quantitative trait loci (eQTLs) – genetic variants linked to specific gene expression levels. We devised eQTL-Capture Hi-C to profile the chromosomal contacts of these loci globally and at high resolution in primary monocytes isolated from 34 donors, and generated chromatin accessibility and gene expression profiles from the same samples. Extending a Bayesian approach that considers both intra- and inter-individual variation, we detected 19 eQTLs linked to distinct promoter contacts, most of which also associated with enhancer accessibility and activity. Capitalising on these shared effects, we next employed a multi-modality Bayesian strategy, identifying hundreds of variants jointly associated with enhancer activity, connectivity, and gene expression. Many of these variants influenced the predicted binding of the architectural protein CTCF and the core myeloid transcription factors GABPA and SPI1; however, they typically did not perturb the canonical binding motifs of these factors. In contrast, one variant associated with *PCK2* promoter contact directly disrupted a CTCF binding motif and impacted the insulation of this promoter from downstream enhancers. Finally, many identified QTLs overlapped with disease susceptibility loci, underscoring the potential role of enhancer-promoter communication in mediating the pathological effects of non-coding genetic variation. Jointly, our findings suggest an inherent functional link between the activity and connectivity of enhancers with relevance for human disease, and highlight the role of genetically-determined chromatin boundaries in gene control.

Introduction

Distal DNA regulatory elements such as transcriptional enhancers play crucial roles in spatiotemporal gene expression control. The binding of transcription factors (TFs) to these elements in response to intrinsic or extrinsic cues promotes a cascade of events that include the recruitment of chromatin remodelers and histone modifiers. Jointly, these events result in the establishment of active chromatin signatures and the engagement of RNA polymerase at these loci. Consequently, the active state of an enhancer is transferred to the promoters of one or more target genes, leading to their increased transcription[1–3].

The transfer of activating signals from distal enhancers to promoters typically (though not always) depends on connections between these elements within the nucleus. These connections can be established through direct 3D chromosomal contacts or potentially through other mechanisms, such as phase separation[3–6]. Previous studies that compared the dynamics of enhancer activation, enhancer-promoter connections and gene expression across cell types and conditions have found that these events are often correlated, although cases whereby enhancer-promoter contacts precede enhancer activation and gene induction are also common[4,7–10]. The precise molecular relationship between these phenomena, however, remains incompletely understood.

Enhancers, and promoter-interacting regions more broadly, are enriched in genetic variants that associate with molecular and cellular traits, such as the risk of common diseases (typically ascertained by genome-wide association studies [GWAS])[11,12] and the expression levels of specific genes ("expression quantitative trait loci", eQTLs)[8,13–16]. Genetic variation across individuals can therefore serve as "nature's perturbation experiment"[17] to provide insights into the mechanisms of enhancer function and enhancer-promoter communication.

In this study, we focused on enhancers that harbour known eQTL variants associated with distal genes in unstimulated human monocytes. We asked whether the transcriptional effects of these variants coincided with their impact on enhancer activity and/or enhancer-promoter communication. To address this question, we performed ATAC-seq and high-resolution Capture Hi-C targeted at eQTL-harbouring regions in monocytes isolated from multiple healthy donors. Adapting two distinct Bayesian methods to identify genetic associations with molecular phenotypes, we detected abundant shared effects of genetic variation at enhancers on promoter contacts, local chromatin accessibility and target gene expression, some of which overlap known disease susceptibility loci. In addition, we identified a distal eQTL variant with opposing effects

on promoter contact and gene expression. We show that this variant disrupts the canonical binding motif of the architectural protein CTCF and likely acts by perturbing the insulation of its target gene promoter from upstream distal enhancers. Jointly, our findings suggest an inherent functional link between the activity and connectivity of enhancers with relevance to human disease and highlight the role of genetically-determined chromatin boundaries in gene expression control.

Results

A compendium of chromatin contacts, open chromatin and gene expression in human primary monocytes isolated from multiple individuals

To profile eQTL-based chromatin contacts at high throughput and resolution, we employed Capture Hi-C (CHi-C) using a frequently-cutting *DpnII* restriction enzyme. We designed a custom capture system targeting 1,197 distal eQTLs in unstimulated monocytes (located at least 10 kb away from their target gene promoters) that were detected in a multi-cohort monocyte eQTL study in 1,480 individuals[18,19] (*Table S1*). Our capture system targeted the lead eQTLs, as well as the variants in tight linkage disequilibrium (LD) with them (R^2 >= 0.9) that mapped within known regulatory regions (see **Methods**). For each eQTL, we aimed to capture the promoter(s) of the associated gene ("eGene") and those of distance-matched "non eGenes", whose expression was not associated with the eQTL. We successfully designed capture probes for 1,074 eGenes and 4,718 control genes, respectively (see **Methods** and *Fig. 1B*). Following sequencing and quality control, the eQTL CHi-C experiment generated a minimum of 9.5 million unique valid on-target reads per donor (median 11.5 million, *Table S2*). By combining the CHi-C data from all 34 donors, we initially obtained a high-coverage consensus dataset consisting of just over 470 million on-target unique reads. Using the CHiCAGO algorithm[20,21], we identified 642,552 significant (score >= 5) contacts at *DpnII*-fragment resolution and 398,464 contacts in bins of 5 kb in this consensus dataset.

To profile chromatin accessibility, we adapted ATAC-seq for crosslinked input material (see **Methods** and *Table S3*) and performed the assay on the same formaldehyde-fixed samples as used for CHi-C. Using a Hidden Markov Model-based pipeline for ATAC-seq[22], we detected 74,373 open chromatin peaks in the ATAC-seq consensus data pooled across individuals. Finally, we profiled gene expression using RNA-seq and genotyped each donor (see **Methods** and *Table S4*).

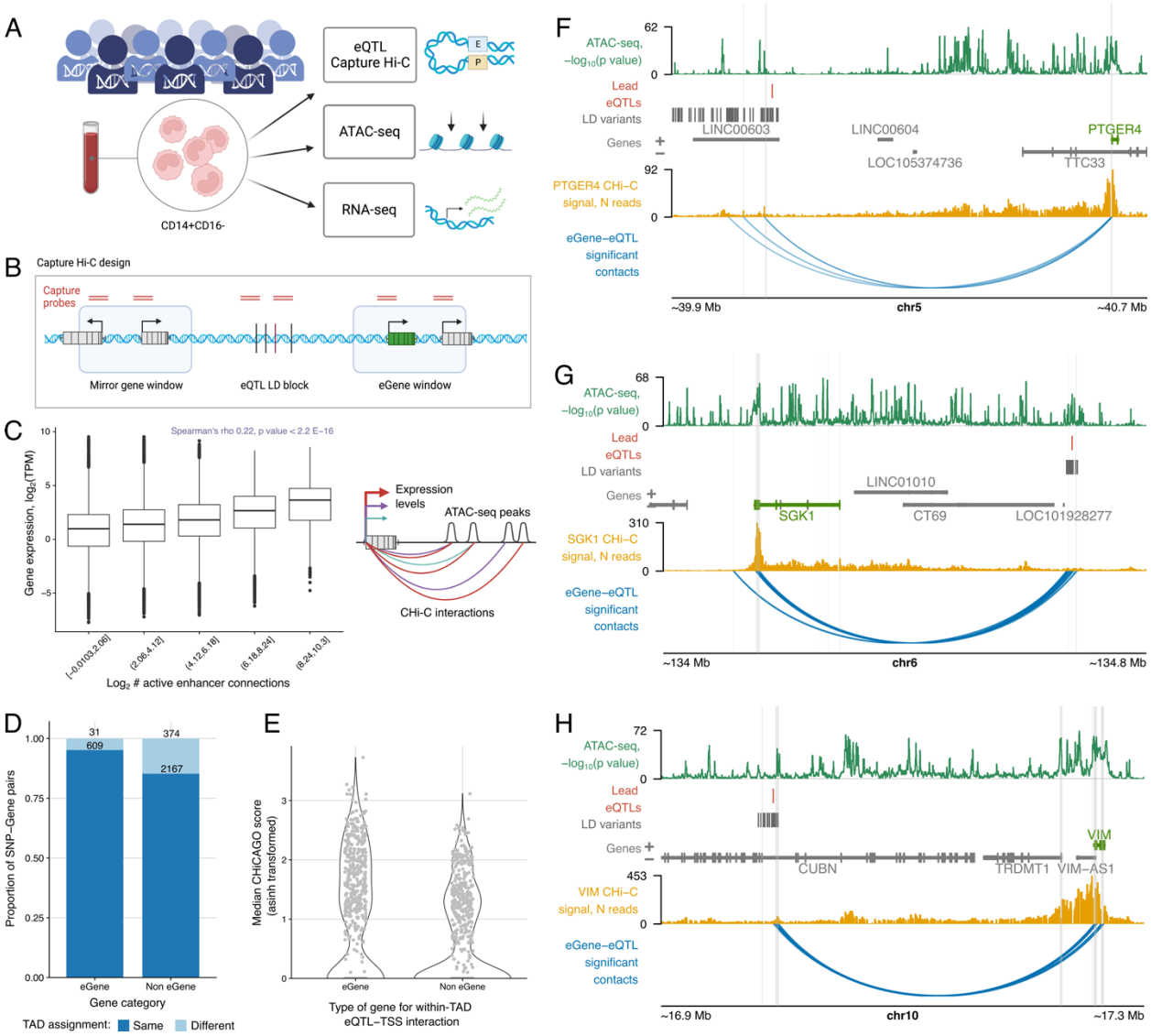


Figure 1. A compendium of eQTL CHi-C contacts and accessibility in primary monocytes*. A) Overview of the main data collection steps. B) Design of the eQTL CHi-C experiment. We designed capture probes targeting DpnII fragments harbouring previously identified lead eQTLs in monocytes. We also included variants in tight LD with the lead eQTLs in regulatory regions, eGene promoters and the promoters of distance-matched "non eGenes", which were similar distances from the eQTLs as the eGenes but not associated with them. C) Relationship between the no. of interacting enhancers and gene expression. D) Degree of TAD sharing between eQTLs and eGenes or eQTLs and non eGenes. E) Inverse hyperbolic sine (ASINH)-transformed median CHiCAGO scores for interactions between eQTLs and eGenes or non eGenes within the same TAD. The score for the eGene is shown against the median score for all captured control genes, per eQTL, including cases where the score was zero. F-H) Examples of eQTLs intersecting ATAC-seq peaks and interacting with the eGenes: PTGER4 (F), SGK1 (G) and VIM (H). ATAC data is presented as -*

log_{10}(p value) for the consensus dataset. eGenes are highlighted in green. Contact profiles show the no. of reads for each other-end fragment contacting the fragment containing the eGene promoter, in the consensus dataset. The shown consensus CHi-C interactions (CHiCAGO score >= 5, DpnII-fragment level) are restricted to those involving the eQTL or a SNP in tight LD and the eGene promoter. Baited regions are shown as a grey highlight. Panels A, B and C were created with BioRender.com.

Using appropriate metrics separately for each assay (CHi-C, ATAC-seq and RNA-seq), we confirmed that overall, the generated data were highly concordant across the cohort (*Fig. S1A-D*). Next, integrating the three assays, we found that as expected, the regions contacted by the captured fragments (containing either eQTL sites or gene promoters) were enriched for open chromatin peaks, as well as for active histone marks detected in other studies in monocytes[23] (*Fig. S1E*). We also observed a positive correlation between the level of gene expression and the number of active regions contacted by a gene's promoter (Spearman's rho = 0.22, p value < 2.2E-16), consistent with previous reports[8,24] (*Fig. 1C*). Jointly, these consensus data provide a high-coverage compendium of chromosomal contacts, chromatin accessibility and gene expression in human primary monocytes.

Distal eQTLs engage in chromosomal contacts with their target genes

We next asked whether eQTLs preferentially contacted their target eGenes compared with distance-matched control genes. We called topologically associating domains (TADs) using Hi-C data from two donors in the cohort. Overall, 95% of eQTLs and their respective eGene promoters shared the same TAD, while a significantly lower proportion of eQTLs shared the same TAD with the promoters of distance-matched control genes (85%, exact Fisher test p value = 3.6E-13) (*Fig. 1D*). Moreover, even within the same TAD, eQTL contacts with eGene promoters tended to have higher CHiCAGO interaction scores than those with control gene promoters (Wilcoxon paired test p value < 2.2E-16; *Fig. 1E*), consistent with previous findings[8,25]. At a CHiCAGO score cutoff of 5, around 70% of eQTLs (816/1,197) had significant chromosomal contacts with the respective eGene promoter, detected either in at least one individual or in the high-coverage consensus CHi-C dataset (as exemplified in *Fig. 1F-H*). As a complementary approach for detecting functional enhancer-promoter contacts, we additionally used the threshold-free "Activity-By-Contact" (ABC) method[26,27], which we recently adapted to CHi-C data (CHi-C ABC)[24]. In this approach, the effects of an enhancer on a target promoter are estimated as the product of local enhancer activity and enhancer-promoter contact frequency. Using ATAC-seq data as a proxy of enhancer activity, ABC predicted functional links between 14,413 putative enhancers and 4,440 genes that were captured in the CHi-C experiment (at an ABC score threshold of 0.012 selected as described in the **Methods**; *Table S5*). However, the ABC-detected pairings only connected 175/1,197

(14.6%) of our targeted eQTLs with their respective eGenes, just 21 of which were not identified by CHiCAGO. The small number of eQTL-promoter contacts detected by ABC is likely due to our focus on distal eQTLs (median eQTL - eGene TSS distance: 57 kb), while ABC preferentially identified short-range promoter contacts (median distance ~23 kb), consistent with previous results obtained with this method in other systems[24,26,27].

Taken together, these findings indicate that eQTLs commonly and preferentially contact the promoters of their target eGenes, suggesting that chromosomal contacts may be involved in mediating the genetic effects of these loci on eGene expression.

Association of expression QTLs with differential promoter contacts

To detect the potential genetic effects of eQTLs on their chromosomal contacts with the eGene promoters from our eQTL CHi-C data, we adapted an allele-specific Bayesian QTL mapping methodology, BaseQTL[28], that considers both within-individual and between-individual allelic effects and explicitly models mappability bias. BaseQTL was originally developed for eQTL calling[28]. Here, we extended the BaseQTL model for calling QTLs based on ATAC-seq and CHi-C read counts (*Fig. 2A*), which included directly accounting for the uncertainty of the QTL genotypes called by imputation (see **Methods**). The density of CHi-C read counts across individuals allowed us to query genetic associations at 1,110 out of the 1,197 targeted eQTL loci, including 974 lead eQTLs and 8,759 proxy variants in tight LD (R^2 >= 0.9).

At a stringent Credible Interval (CI) threshold of 99%, we detected 19 significant variants associated with differential contacts between 13 eQTL-containing *DpnII* fragments and 9 eGene promoters, which we termed "contact eQTLs" (*Fig. 2B* and *Table S6*). The distances between the contact eQTLs and the respective eGene promoter regions ranged from ~32.5 to ~283 kb, with a median of 72.3 kb.

The thrombospondin 1 (*THBS1*) locus had the largest number of contact eQTLs (seven, in tight LD with each other [R^2 >= 0.9]). The alternative alleles of these variants were associated with stronger contacts with two *THBS1* promoters located ~250 kb away from them (*Fig. 2C*). In the *C7orf50* and *NAAA* loci, we also detected multiple contact eQTLs in tight LD with each other, whilst only one contact eQTL was detected within each of the remaining six loci (*NFE2L3*, *PTPRE*, *SVIL*, *KCNK13*, *PCK2* and *TFPT*) (*Fig. 2D*). Notably, the *TFPT* contact eQTL (rs10415777) was also in tight LD with an eQTL (rs35154518) for another gene, *OSCAR*. The expression levels of *TFPT* and *OSCAR* were correlated across the donors in our cohort (Pearson's correlation r = 0.54, p value = 0.001), suggesting a shared genetic control of these two genes. Moreover, rs10415777 showed

the same direction of effect for contacts with both *TFPT* and *OSCAR* (*Fig. 2E*). Although the effect of rs10415777 on *OSCAR* did not pass our stringent significance threshold, these observations raise the possibility that rs10415777 may affect the promoter contacts of the regulatory region encompassing this variant with the promoters of both genes.

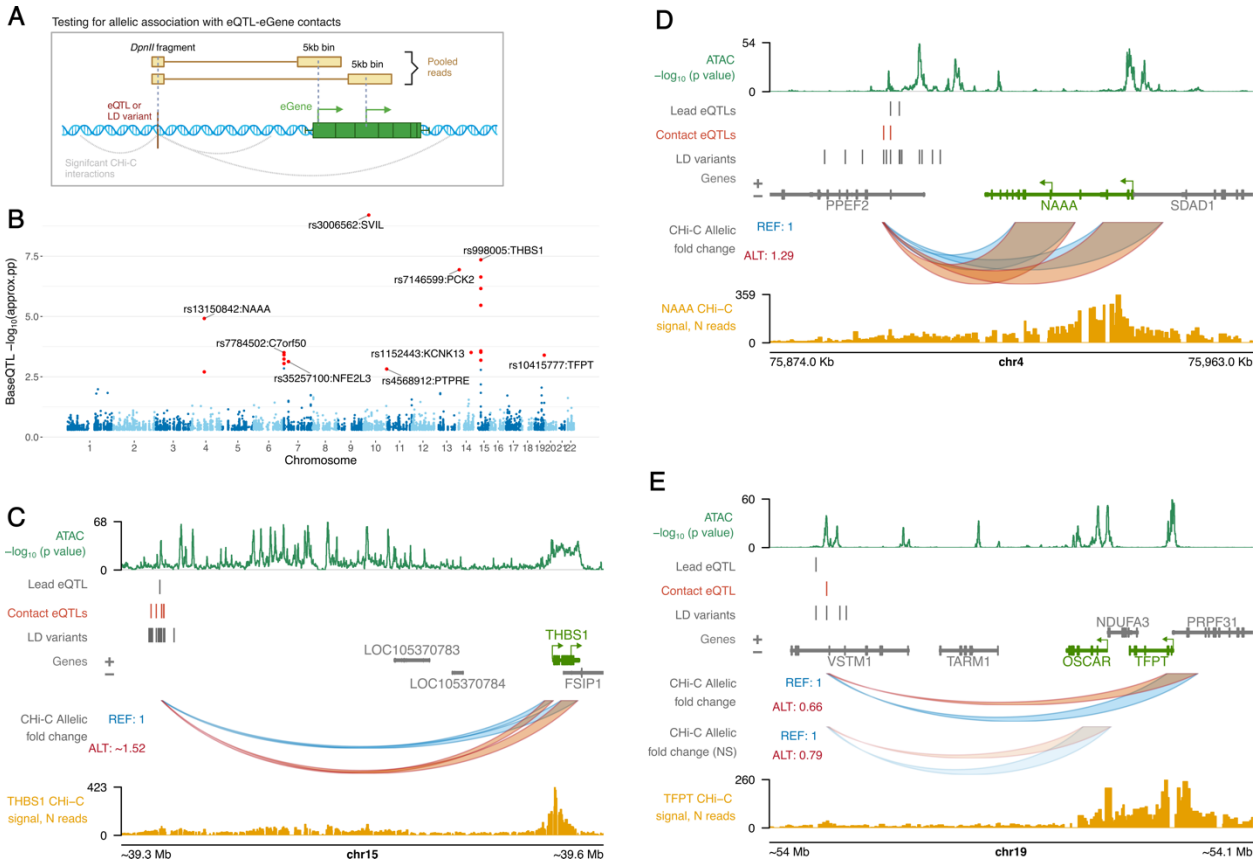


Figure 2. Detection and examples of contact eQTLs*. A) Strategy for testing eQTL-eGene contacts. B) Manhattan plot of BaseQTL results. The y-axis shows the negative logarithm of the probability that the candidate SNP is not a true QTL for promoter contact: approximate posterior probability (approx.pp; see Methods). The significant contact eQTLs (0 excluded from the 99% credible interval) are depicted in red. The contact eQTL with the highest probability of being a true QTL for promoter contact is labelled in loci with multiple significant contact eQTLs. C-E) Genomic visualisation of significant contact eQTLs in the THBS1, NAAA and TFPT loci. The arches, whose heights correspond to the allelic fold change in contact, show the tested contacts between the eQTL and promoter(s) of the eGene (one eQTL-containing DpnII bait fragment is shown in each case). CHi-C read counts at DpnII fragment resolution are shown from the viewpoint of the eGene promoter (yellow signal tracks). ATAC data is presented as -log$_{10}$ (p value) for the consensus dataset. Panel A was created with BioRender.com.*

We asked whether eQTLs were more likely to influence promoter contacts compared with variants that contacted promoters in 3D, but did not associate with gene expression. To test this, we considered a set of 11,582 control ('non-eQTL') variants with significant evidence of promoter contact in our consensus CHi-C dataset (CHiCAGO score >= 5) that are not known to be eQTLs in monocytes[19] or across blood cells[18] and assessed their association with the promoter contact strength using BaseQTL (*Table S6*). Despite the 1.4-fold larger number of non-eQTL variants compared with that of the eQTLs included in the analysis, only two non-eQTLs, mapping to a single genomic locus, showed association with contact strength (rs7905438 and rs7917620 contacting the promoter of *RNU6-6P* from a distance of 1 kb). We confirmed the significant enrichment of genetic effects on promoter contact among eQTLs compared with promoter-interacting non-eQTLs using a permutation test that compared the respective posterior probabilities of association, accounting for read coverage (p value < 1E-4; see **Methods**).

Most contact eQTLs associate with local chromatin activity

To probe the relationship between the genetic effects of enhancer connectivity and activity on gene expression, we additionally detected QTLs associated with differential ATAC-seq signals in the same individuals (ATAC-QTLs; *Fig. 3A* and *Table S7*), as well as validated the effects of multi-cohort eQTLs using the RNA-seq data from our cohort (*Fig. S2* and *Table S8;* see **Methods** for details). Genome-wide, we detected 34,900 ATAC-QTL variants within a 5 kb vicinity of 6,855 ATAC-seq peaks. ATAC-QTLs directly overlapped 104/1,197 (~9%) lead eQTLs profiled in the eQTL-CHiC experiment and were in tight LD (R^2 >= 0.9) with an additional 119 lead eQTLs. ATAC-QTLs also directly overlapped 10/19 contact eQTLs (14 accounting for LD). Notably, the allelic effects of these variants on both contact strength and enhancer accessibility were exclusively in the same direction (*Fig. 3B*). In all but one case, these effects were also in the same direction as gene expression (*Fig. 3C*).

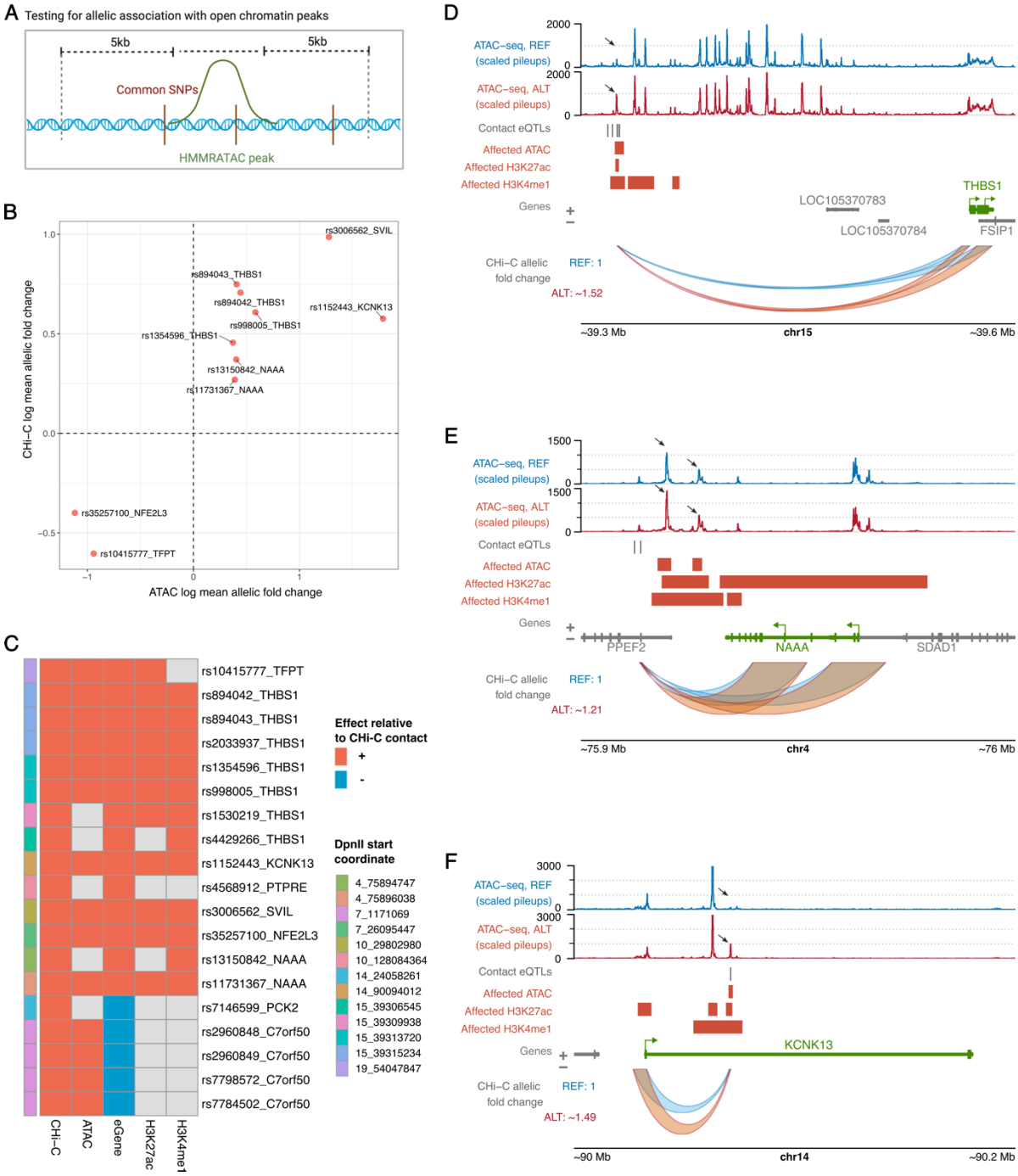


Figure 3. Shared genetic effects on enhancer activity, enhancer-promoter contact and gene expression. *A) Strategy for detecting ATAC QTLs with BaseQTL. B) Shared allelic effects of QTLs on accessibility and promoter contact at 99% credible interval (CI). C) Heatmap of cross-trait effects for contact eQTLs. The positive effect allele (either REF or ALT) is shown for CHi-C and the direction of effects in other traits (whether log aFC or beta) are shown relative to this. We only show effects for features within 5 kb of the contact eQTL, and we accounted for LD (R^2 >= 0.9). The ATAC QTL effects were from BaseQTL analysis*

within our 34-donor cohort, whereas the remainder of effects (gene expression, H3K27ac and H3K4me1) were curated from outside of our 34-individual cohort. The eQTL effects were taken from the original monocyte study or from Blueprint. The histone modifications H3K27ac and H3K4me1 were taken from Blueprint WP10 Phase 2. D-F) Epigenetic mechanisms within example contact eQTL loci: D) THBS1, E) NAAA, F) KCNK13. The red regions show upregulated peaks of ATAC, H3K27ac and H3K4me1 associated with the contact eQTL (or SNPs in LD, R^2 >= 0.9); the peaks shown in these plots were not restricted to 5 kb from the contact eQTL. ATAC-seq tracks show the pileups for merged homozygous reference (blue) or homozygous alternative (red) donors for one of the contact eQTLs in each locus. Arrows denote the ATAC-seq peaks that were also associated with the contact eQTLs. Panel A was created with BioRender.com.

To further assess the genetic effects of contact eQTLs on enhancer activity, we integrated QTL data for histone modifications in human monocytes from the Blueprint WP10 Phase II release[29]. We found that 53% (10/19) of the contact eQTLs overlapped QTLs for nearby active histone marks H3K27ac and H3K4me1 that were within 5 kb of the contact QTL itself (accounting for LD [R^2 >= 0.9]) (*Fig. 3C*). The direction of effect for these histone mark QTLs was always consistent with that of contact, accessibility and eGene expression (*Fig. 3C*). For instance, all contact eQTLs in the *THBS1* locus were also QTLs for one or more local features of activity (ATAC-seq, H3K27ac and H3K4me1), and the signals for all these features were stronger in the alternative allele (*Fig. 3D*). We observed a similar pattern of stronger enhancer activity signals associated with the alternative allele of the single contact eQTL in the *SVIL* locus (*Fig. 3E*). In addition to their effects on local enhancer activity, we also found that some of the contact eQTLs were associated with increased chromatin accessibility and active histone marks at the promoters of their respective eGenes. For example, contact eQTLs for *NAAA* and *KCNK13* were associated with H3K27ac levels both proximally to these variants and at the promoters of these genes (*Figs. 3E* and *3F*, respectively).

In conclusion, contact eQTLs typically, but not always, also associated with differential chromatin activity *in cis*, suggesting abundant shared genetic effects on enhancer activity, connectivity and gene expression.

Joint detection of shared genetic effects on enhancer accessibility, connectivity and gene expression

The approach described above identified genetic associations with enhancer-promoter contact and showed overlap with genetic effects on other molecular phenotypes ("modalities"), despite testing each one separately. Given this overlap and our relatively small sample size, we sought to

increase the detection power by jointly modelling all three modalities (gene expression, chromatin accessibility and eQTL-promoter contact). To detect such "trimodal QTL" variants, we adapted a Bayesian QTL mapping framework, GUESS[30,31], developed to enhance the power to detect joint QTLs by leveraging the information contained in multiple molecular traits. Specifically, we defined 5,564 windows containing 24,485 genotyped variants within ~5 kb proximity to the tested distal eQTL-eGene contacts and ATAC-seq peaks (*Fig. 4A*). Within each window, we searched for models consisting of one or a combination of genetic variants that jointly explained the three phenotypes. At 5% FDR[32] in each window and following post-hoc filtering (see **Methods**), we identified 877 partially-overlapping windows containing 614 significant trimodal QTLs (*Fig. 4B*; *Table S9*). In total, trimodal QTLs were associated with 701 eQTL-eGene contacts, 298 open chromatin peaks and 218 eGenes (*Fig. 4B*).

The GUESS approach replicated the effects of 79% (15/19) of the CHi-C BaseQTL contact eQTLs on the contacts with six respective eGene promoters (*THBS1*, *NAAA*, *SVIL* and *KCNK13, TFPT* and *C7orf60*), either via direct variant overlap (N = 5) or by overlap with other variants within the considered GUESS window and in tight LD ($R^2 > 0.9$) with the lead trimodal QTL (N = 10). Overall, the estimated variant effects for each of the three modalities (contact, accessibility and expression) were highly correlated between the BaseQTL and GUESS approaches, while the significance of the effects was higher for GUESS, indicative of the increased power of this approach (*Fig. S4*). Consistent with this, GUESS detected many novel genetic associations with a shared effect on contact, accessibility and expression. An example of a novel shared effect was the trimodal QTL rs61837569 at the Toll-like receptor 5 (*TLR5*) locus (*Fig. 4C)*. This variant was in tight LD with the lead *TLR5* eQTL from the original monocyte study ($R^2 >= 0.995$) and was selected by GUESS out of a set of 10 variants in the window (*Fig. 4C*). In a small number of other loci, GUESS identified multiple trimodal QTLs that jointly explained the observed genetic effects on contact, accessibility and gene expression (*Fig. 4D*). For example, in the Growth Factor Receptor Bound Protein 2 (*GRB2*) locus, GUESS detected two trimodal QTLs (rs12936583 and rs62090276), which were in moderate LD with each other (R^2 ~0.59) and selected out of 13 possible variants (*Fig. 4E*).

In summary, by jointly analysing functionally-related molecular traits, we have identified a large set of monocyte QTLs with shared effects on chromatin activity, enhancer-promoter communication and gene expression, which we release as a resource in *Table S9*.

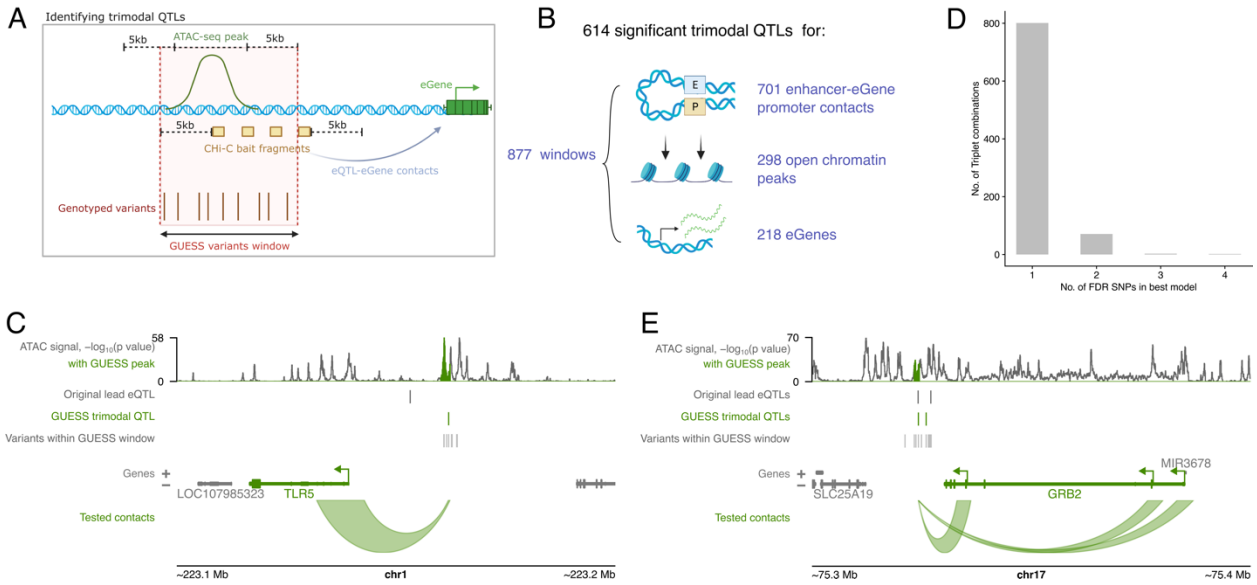


Figure 4. Detection and examples of trimodal QTLs*. A) Strategy for detecting trimodal QTLs using GUESS. Regions within 5 kb of ATAC-seq peaks and CHi-C DpnII bait fragments were identified and all genotyped variants were queried within these regions. B) Overview of the significant findings from the GUESS analysis. C) An example of a GUESS locus where the best model (combination of genetic variants with the largest marginal likelihood score) contained a single trimodal QTL (also significant at 5% FDR) that was associated with contact with the eGene (TLR5), chromatin accessibility (highlighted in green in ATAC-seq track) and expression of TLR5. D) No. of significant QTLs in each of the GUESS best models. E) An example of a GUESS locus where the best model contained two trimodal QTLs (also significant at 5% FDR) that were both associated with contact with the eGene (GRB2), chromatin accessibility (highlighted in green in ATAC-seq track) and expression of GRB2. Panel A was made with BioRender.com.*

Genetic effects on enhancer-promoter contacts are likely mediated by a diverse range of transcription factors

To determine the potential mechanisms underlying the phenotypic effects of the identified QTLs, we searched for evidence of protein binding at these loci. For this analysis, we combined contact eQTLs and trimodal QTLs, obtaining 627 distinct variants, which we will refer to collectively as "contact QTLs" (cQTLs). First, we assessed the intersection of cQTLs with the *in vivo*-detected protein-binding regions. For this, we queried cQTL locations in the ReMap catalogue, a database of uniformly processed chromatin immunoprecipitation-sequencing (ChIP-seq) data for 1,210 distinct proteins[33]. As background, we used the locations of all 25,151 eQTL variants tested in the

BaseQTL and GUESS analyses. We found that cQTLs were significantly enriched for the binding of seven TFs and cofactors compared with all eQTLs (*Fig. 5A*). These included TFs with known roles in the innate immune response (RXR, EGR1, STAT1, ETS1), histone modifying enzymes (the H3K9 demethylase JMJD1C and the H3K4 methyltransferase KMT2B/MLL2) and the androgen receptor (AR), known to be important for monocyte development and proliferation[34]. Moreover, despite the diverse range of cell types (N=739) available in the catalogue and included in this analysis, all of the seven ChIP-seq datasets showing enrichment at cQTLs had been profiled in monocytes or related cell types (*Fig. 5A* and *Table S10*). Overall, this result is consistent with the potentially causal role of the cQTLs in modulating enhancer function in monocytes.

We next examined the potential allelic impact of cQTL variants on TF binding to enhancers. For this, we took advantage of a deep learning framework, Enformer, which was pre-trained on a multitude of public genomic datasets, enabling us to infer the effects of sequence variation on signals in these training data. In this analysis, we focused on ChIP-seq data for sequence-specific TFs, with a total of 1433 ChIP-seq tracks for 662 TFs available in the Enformer training set (*Table S11*). We used Enformer to predict the effects of cQTL variants on ChIP signals in the vicinity of these variants (128 bp each side of the cQTL variant), determining the significance of predicted perturbations against a random background (see **Methods**). At 5% FDR (Benjamini-Hochberg), approximately ~12.5% of cQTL variants (78/627, associated with 71 genes) were predicted to significantly impact the ChIP-seq signals for at least one TF (*Fig. 5B*).

Overall, Enformer predicted that the binding of just over half of the examined TFs (373/662) was perturbed by at least one cQTL (*Fig. 5B)*. Consistent with this, the ATAC-seq footprints of many of these TFs (64 out of 113 tested, 57%) were significantly enriched in the 128-bp vicinity of cQTLs (5% FDR; *Fig S3A*). Notably, over two thirds of TF-perturbing cQTLs (55/78, 70.5%) affected the predicted binding of multiple TFs, with a median of three TFs per variant (*Fig 5B*), which is consistent with abundant TF cooperativity at enhancers[35–37]. To identify groups of TFs typically perturbed by the same cQTLs, we used hierarchical clustering, defining 20 coherent clusters that ranged in size from one to 74 TFs (colour-coded in *Fig. 5C* for the top perturbed TFs; *Figs. 5D and S3B*). These clusters brought together TF isoforms (such as CEBPB and CEBPG; SP2 and SP3), but also highlighted the predicted joint variant effects on the binding of such factors as JUN and FOS that form a heterodimer on the DNA (known as activator protein 1 [AP-1]); the TATA-binding protein [TBP] and ETS-family TFs ELF1 and GABPB1; Ikaros-family protein IKZF1 and activating transcription factor 3 [ATF3] (*Fig. S3B*).

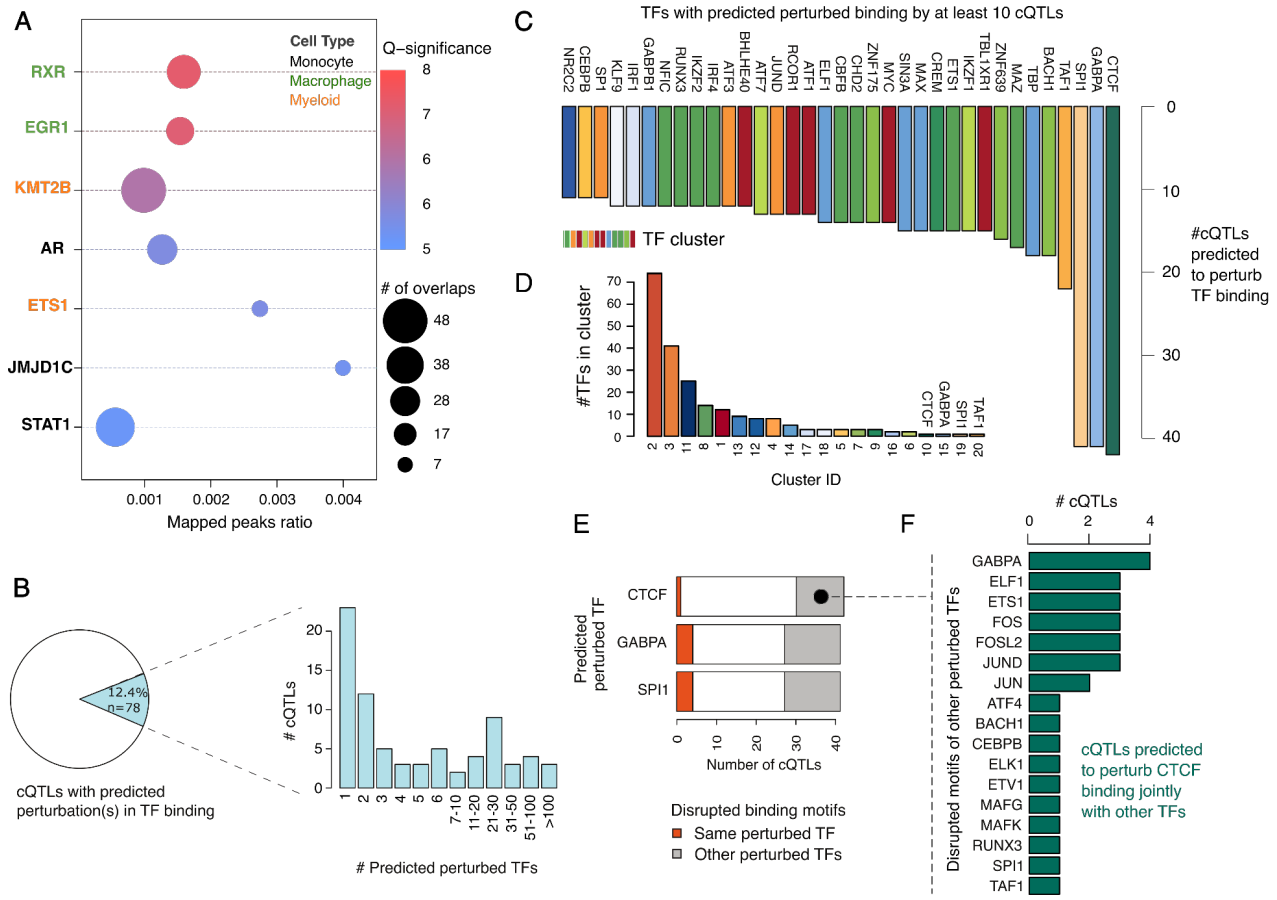


Figure 5. Effects of contact QTLs on TF binding. *A) Significantly enriched TFs at cQTLs from the ReMap catalogue (5% FDR, Benjamini Hochberg). The cell types for these factors included monocytes (black text) macrophages (green text) and myeloid (orange). B) Pie chart: Proportion of cQTLs with predicted perturbations in TF binding, detected by Enformer. Bar chart: Histogram of the no. of TFs predicted to be perturbed by cQTLs with at least one TF perturbation. C) TFs whose binding was predicted to be perturbed by at least 10 cQTLs. The colours mark TF perturbation clusters, shown in full in Fig. S3. D) No.s of TFs by perturbation cluster. TFs forming part of single-TF clusters are labelled on the plot. E) No.s of cQTLs with predicted effects on the binding of GABPA, SPI1 and CTCF that disrupted the known sequence binding motif for either the same (red) or other predicted perturbed TFs (grey). F) TF motifs disrupted by cQTLs that were predicted by Enformer to perturb CTCF binding jointly with other TFs, but did not disrupt the known CTCF motif.*

The predicted binding of the architectural protein CTCF and the ETS-family TFs GAPBA and SPI1 (also known as PU.1) were most affected by cQTLs. Overall, ~10% of cQTLs (63/627), including ~80% (63/78) of those that were predicted to perturb the binding of at least one TF, affected the predicted binding of CTCF, GABPA and/or SPI1, either alone or in combination with other TFs. Our partitioning placed each of these three TFs into their own single-TF cluster (*Fig. 5D and S3B*),

indicating that the patterns of their predicted binding perturbations by cQTLs were generally distinct from one another and from the other TFs included in the analysis. Many cQTLs perturbing the predicted binding of CTCF, GABPA or SPI1 localised in the 128-bp vicinity of the canonical sequence recognition motifs or known ChIP peaks of the respective factor (*Fig S3C*). Notably however, only a small minority of cQTLs directly disrupted the recognition motifs of these factors, while many more of them disrupted the recognition motifs of other TFs whose binding they were also predicted to perturb (*Fig. 5E* and *Table S11*). For example, while 26/42 of the cQTLs predicted by Enformer to perturb CTCF binding localised nearby a CTCF motif or ChIP peak (*Fig S3C*), only one cQTL disrupted the CTCF motif (*Fig. 5E*). At the same time, 12 of these cQTLs disrupted the binding motifs of other TFs whose binding they were also predicted to perturb, including ETS family TFs GABPA, ETS1 and ELF1; as well as JUN/FOS and others (*Fig. 5F*). cQTLs with predicted GABPA and SPI1 binding perturbations showed a similar trend (*Fig. S3D*). Interestingly, those cQTLs that did affect the cognate motif could also perturb an overlapping motif for another factor. For example, the alternative allele of cQTL rs881382, a trimodal QTL for *IGFBP7*, disrupted the binding motif for SPI1, while improving the motif match for another core monocyte TF, IRF8 (*Fig. S3E*). Jointly, these results suggest that the effects of cQTLs on TF binding can be cooperative or indirect, potentially mediated by perturbations in the binding sites of either known or not yet identified binding partners of these TFs.

Finally, we asked whether cQTLs perturbed the binding of a different set of TFs compared with variants associated with enhancer activity, but not promoter contacts. To this end, we additionally ran Enformer predictions for the 513 ATAC-eQTLs identified in the BaseQTL analysis that were not detected as significant contact eQTLs (BaseQTL, 95% CI) and as trimodal QTLs (GUESS, FDR 5%). We then used logistic LASSO regression to identify TF binding perturbations that can discriminate between cQTLs and ATAC-eQTLs. At an optimal level of shrinkage determined by cross-validation (see **Methods**), the LASSO models incorporated the binding perturbations of several TFs, including ZNF639, MYC, IRF1, KLF9 and GAPBA that were preferentially observed at cQTLs, and those of EGR2, MEF2A, SRF and others that were enriched at ATAC-eQTLs (*Fig. S3F*). However, overall the power of these LASSO models to discriminate between the two QTL classes remained low (AUC=0.59), and an even lower power was achieved with random forest classification (AUC=0.52; not shown). Therefore, trimodal QTLs could not be straightforwardly discriminated from ATAC-eQTLs based on the predicted patterns of perturbed TF binding alone. This suggests that factors other than the identity of perturbed TFs may determine whether genetic variants have either joint or independent effects on enhancer activity and enhancer-promoter contact.

A contact eQTL affects *PCK2* expression by modulating CTCF binding and chromatin insulation

Contrary to the majority of BaseQTL-detected contact eQTLs that showed a consistent direction of allelic effect on gene expression and promoter contact, variants identified in the *PCK2* and *C7orf50* loci showed opposite direction of effects on these two properties (*Fig. 3C*). Notably, neither of these variants associated with "active" histone marks, and the *PCK2* contact eQTL variant also did not associate with chromatin accessibility. To determine the mechanism of these variants, we turned to the Enformer predictions of perturbed TF binding by these variants. Whilst none of the four *C7orf50* contact eQTLs were predicted to perturb any TFs, the *PCK2* contact eQTL variant rs7146599 was predicted to strongly affect CTCF binding (FDR < 0.01%). Moreover, this variant was the only cQTL to perturb the canonical motif of CTCF (*Fig. 6A*). Consistent with these observations, the alternative allele of this cQTL was associated with a weaker contact with the *PCK2* gene promoter (*Fig. 6B*).

In contrast to the decreased contact with *PCK2* promoter and predicted decreased CTCF binding, the alternative allele of rs7146599 is associated with increased *PCK2* expression across multiple eQTL studies in monocytes[19,29,38] and whole blood[13,18,39,40] (*Fig. 6C*). While we did not find a significant association between rs7146599 and *PCK2* expression in our own cohort, we observed the same direction of effect (log allelic fold change = 0.075). We therefore asked whether the *PCK2* promoter contacted additional enhancers in the alternative genotype. To address this, we compared the patterns of *PCK2* promoter interacting regions between donors that were homozygous for the reference or alternative allele of rs7146599. In the homozygous alternative genotype, we observed increased interactions between the *PCK2* promoter and upstream regions (CHiCAGO score >= 5, 5 kb resolution; highlighted in red in *Fig. 6B*) containing open chromatin and predicted enhancer elements in monocytes (based on data from Ensembl regulatory build[41]). We therefore hypothesise that rs7146599 perturbs the function of a CTCF-dependent insulator element that shields *PCK2* from upstream enhancers (*Fig. 6D*). Notably, both rs7146599 and the *PCK2* promoter shared the same TAD and neither of them is located at a TAD boundary based on our monocyte Hi-C data, therefore the observed effects likely take place at a sub-TAD level.

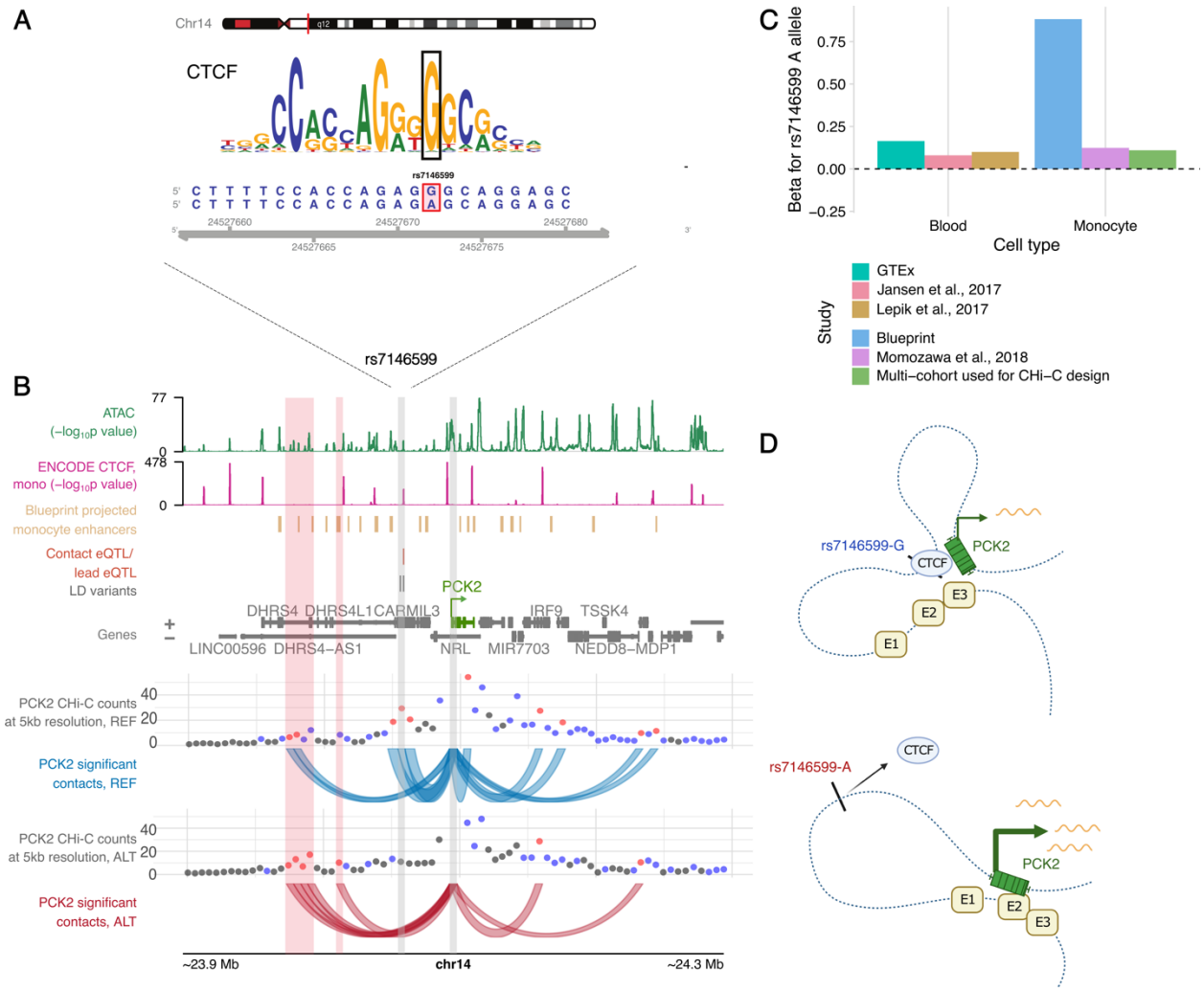


Figure 6. A contact eQTL perturbs CTCF binding and chromatin insulation. *A) Position of rs7146599 within the CTCF motif. B) Effect of genotype on contact profiles in the locus. Individuals were split into monozygous REF (N = 11) or ALT (N = 9) genotypes and merged in CHiCAGO to produce an average count no. in bins of 5 kb. Contact profiles are shown from the viewpoint of the PCK2 promoter, with the location of the promoter and the contact QTL highlighted with grey rectangles. Maintained/gained interactions upstream of the PCK2 promoter in the alternative genotype are highlighted with red rectangles. C) eQTL effects of rs7146599/PCK2 across multiple cohorts. Betas with respect to the alternative allele (A) are shown in studies of whole blood (GTEx[42], Jansen et al., 2017[39] and Lepik et al., 2017[40]) and monocytes (Blueprint[29], Momozawa et al., 2018[38] and the monocyte multi-cohort analysis used to design the eQTL CHi-C experiment[19]). D) Proposed mechanism schematic showing how perturbed CTCF binding at rs7146599 could affect the insulation of the PCK2 promoters from distal enhancers, denoted as "E". Panel D was made with BioRender.com*

Contact QTLs associate with healthy and pathological genetic traits

As a first indication of the effects of cQTLs on physiological traits and disease, we examined the representation of these variants in the GWAS catalog[43]. Overall, 11.6% of cQTLs (73 out of 627, including 3 out of 19 contact eQTLs) overlapped the reported GWAS variants for a total of 91 traits accounting for LD ($R^2 > 0.8$; *Table S12*), with 52 cQTLs directly coinciding with these reported GWAS variants. The associated traits included several in which monocytes are a likely causal cell type, such as haematological and inflammatory conditions. For example, four cQTLs were associated with white blood cell count, implicating the target eGenes nuclear receptor corepressor 1 (*NCOR1*, located ~200 kb away from the trimodal QTL rs9910148) proposed as a key immunometabolic regulator[44] (*Fig. 7A*), the lysosomal gene *LAMP1* associated with mononuclear phagocyte activation[45] (~34 kb away from the trimodal QTL rs9604045), as well as peptidylglycine alpha-amidating monooxygenase (*PAM*, ~166 kb away from the trimodal QTL rs2431320) and EP300-interacting inhibitor of differentiation 2 (*EID2*, ~96 kb away from the trimodal QTL rs1865092).

The *LAMP1*-associated rs9604045 and another trimodal QTL (rs7488791, associated with the expression of a scavenger receptor *SCARB1,* located ~33 kb away) were also GWAS variants for the levels of C-reactive protein, a well-established biomarker of inflammation[46]. In addition, two trimodal QTLs (rs4698412 and rs11724635) were associated with Parkinson's disease, which is notable given the emerging role of the innate immune system in this disorder[47]. Both Parkinson's disease-associated variants shared the same target eGene, bone marrow stromal cell antigen 1 (*BST1*, 21-28 kb away) that has a known role in humoral inflammatory response[48] (*Fig. 7B*). As a final example, two trimodal QTLs (rs10093709 and rs10112966) overlapped GWAS variants associated with macrophage migration inhibitory factor levels and were eQTLs for nicotinate phosphoribosyltransferase expression (*NAPRT*, located ~25 kb away), a critical metabolic regulator of inflammation[49].

Jointly, these examples point to the relevance of enhancer-promoter contacts in mediating the effects of genetic variants associated with healthy and pathological traits.

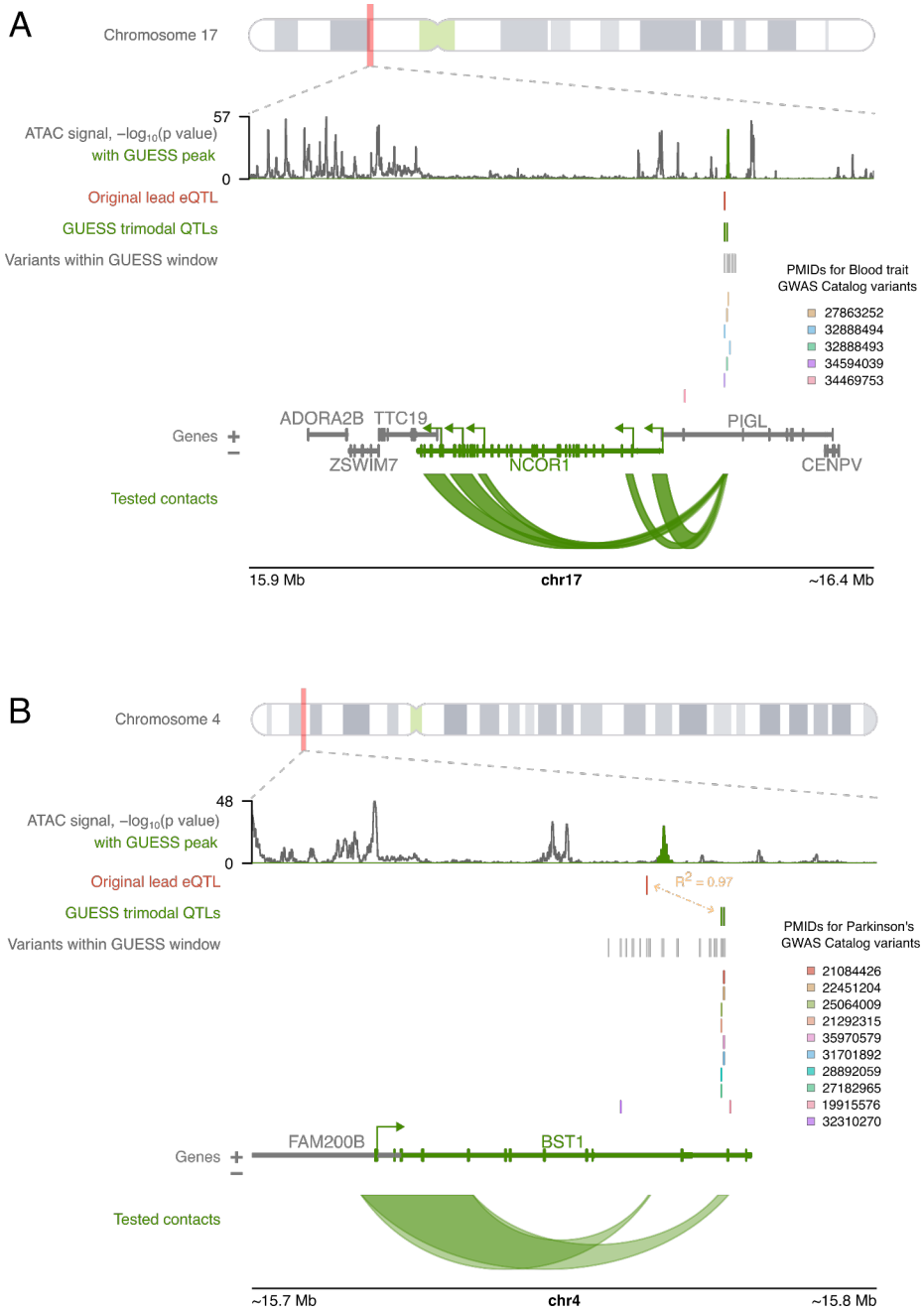


Figure 7. Trimodal QTLs overlap variants associated with healthy and pathological traits. *The intersection between trimodal QTLs and GWAS variants is presented in two loci: A) a blood cell traits locus at NCOR1 and B) a Parkinson's disease locus at BST1. To generate these plots, an intersection was performed between the trimodal QTLs and the GWAS Catalog (which contains the top susceptibility variant for each independent locus in each study). Next, the GWAS Catalog was filtered for the overlapping traits ("Parkinson's Disease"*[50–60] *in A and "Neutrophil count"*[50,61–63] *or "Lymphocyte percentage of white blood cells"*[50,64] *in B) and all variants within the Catalog were plotted in each locus.*

Discussion

In this study, we have used multimodal profiling of molecular phenotypes across individuals to establish the effects of genetic variants at enhancers on chromatin activity, connectivity and gene expression in human primary monocytes. We observed widespread shared genetic effects on these three modalities, suggesting a common underpinning molecular mechanism. However, we also found an example whereby a distal variant likely influences gene expression through perturbing a CTCF-dependent chromatin insulator and the resulting co-option of downstream enhancers.

Assessing the effects of genetic variants on chromosomal contacts by multi-individual chromosome conformation capture analyses presents inherent challenges, particularly in a high throughput setting and for longer-range contacts. The large numbers of sequencing reads per donor required to resolve individual contacts and the complexity of the experimental protocols hamper the statistical power of this analysis, in terms of both feasible sequencing coverage and cohort size. As a result, only limited numbers of contact QTLs were identified using Hi-C-based analyses in human cell lines[65,66]. We and others have previously used Promoter Capture Hi-C in small-scale analyses of allelic effects on enhancer-promoter contacts[25,67]. Here, we targeted the Capture Hi-C system to short restriction fragments containing eQTL variants in addition to their target eGene promoters, which ensured that the sequencing reads directly encompassed the variants of interest. This has enabled us to conduct within-individual allelic analysis for heterozygous donors in addition to genotype-based QTL detection across all individuals. As a result, with a relatively small cohort size of 34 donors, we were able to detect 19 contact eQTLs in a single-modality analysis, as well as a multitude of ATAC eQTLs. The predominantly shared genetic effects on enhancer activity, connectivity and gene expression revealed by this analysis have then enabled the joint detection of genetic effects on these modalities. This multi-modality Bayesian approach has provided a boost in statistical power, identifying over 600 putative "trimodal" QTLs, mirroring the well-known advantages of multivariate analysis of variance (MANOVA) for detecting patterns between multiple correlated dependent outcomes over the univariate analysis (ANOVA) that tests one outcome at a time[68].

Our choice of eQTLs as anchors for this analysis was also motivated by the fact that enhancers can efficiently buffer the effects of genetic variation, even when this variation results in detectable effects on transcription factor binding and chromatin state[69,70]. Furthermore, genes are commonly controlled by multiple enhancers with a considerable degree of redundancy[71,72]. Consequently, disruptions in enhancer activity may not lead to prominent alterations in target

gene expression, especially in stable conditions[73,74]. Therefore, we focused on variants that have demonstrated effects on gene expression in unstimulated primary monocytes to prioritise cases where the observed associations between these variants and enhancer activity and connectivity are more likely to be functionally meaningful. Consistent with this expectation, a significant minority of the identified QTLs have known associations with diverse genetic traits and diseases, including those directly implicating monocytes.

We have identified hundreds of putative trimodal QTLs with shared effects on enhancer activity, enhancer-promoter connectivity and gene expression, and given the limitations of power, particularly in the Capture Hi-C modality, it is likely that the true number of variants exhibiting such effects is even larger. Computational predictions suggest that many trimodal QTLs affect the binding of one or more TFs, including most prominently the looping factor CTCF[75] and the transcription factors GABPA and SPI1/PU.1, both of which were implicated in enhancer-promoter looping[67,76,77]. In particular, CTCF can stabilise chromosomal loops, including those formed between promoters and enhancers, by constraining the cohesin complex on the chromatin[75,78–80]. As a result, the predicted perturbations in CTCF binding by cQTLs could provide an explanation for some of the observed genetic effects on enhancer contacts. Notably, with the exception of one special case, the canonical sequence recognition motif of CTCF remained unaffected by the cQTLs predicted to perturb CTCF binding. This suggests that, provided the computational predictions of CTCF binding perturbations by these variants are accurate, these effects could be indirectly induced by cQTLs, possibly through their broader impact on enhancer activity, which may affect the dynamic recruitment of CTCF to these regions[81–83]. Considered jointly with the diversity of factors whose binding could be affected by trimodal QTLs, these results point towards an inherent link between enhancer activity and connectivity that may be largely independent of the identity of specific TFs recruited to enhancers. Consistent with this notion, we did not observe a strong predictive signature of specific TF binding perturbations at contact QTLs that could distinguish them from the “non-contact” ATAC eQTLs. Therefore, the ability of genetic variants to affect enhancer-promoter contact could be determined by the additional parameters of enhancers or their broader genomic context[84] that are independent of these variants themselves and their impact on TF binding.

Which molecular mechanisms could link enhancer activity and connectivity? The cohesin complex, and particularly its isoform containing the STAG2 subinit, is known to be recruited to active enhancers, including those that do not bind CTCF[85,86]. This recruitment could be directly impacted by genetic aberrations in enhancer activity. Mechanisms of cohesin recruitment to active chromatin are not fully understood, but may involve interaction with the H3K4me1 histone

mark[87] or with acetylated histones[88]. However, a significant number of enhancer-promoter contacts, particularly those in the shorter distance range, are likely independent of both cohesin and CTCF[89–92]. It is possible that in these cases enhancer-promoter proximity is facilitated by phenomena such as liquid-liquid phase separation[6,93–95] that can be directly mediated by the recruited TFs[96–98] and core cofactors, including the Mediator complex[97,99–102]. Finally, transcription itself has a known role in chromatin looping[103,104]. While it is likely that other mechanisms are required for the initial enhancer-promoter communication upon gene induction, recent evidence suggests a role for RNAP II and transcription in this process[105,106].

Finally, we show that in the *PCK2* locus, direct disruption of the canonical CTCF sequence recognition motif by a genetic variant leads to perturbed promoter contact and increased distal gene expression without an apparent effect on chromatin activity *in cis*. The influence of genetic perturbations of CTCF binding on 3D chromosomal architecture is expected and has been reported previously[66,107]. It is also known that disruption of enhancer-promoter contacts, including through perturbations of architectural proteins, can drive down gene expression[89–91]. However, in the specific case we have identified, genetic perturbation of a CTCF binding site and ablation of its contact with *PCK2* promoter instead led to increased expression of *PCK2*. We propose that this effect can be attributed to the potential co-option of enhancers located further upstream of the *PCK2* promoter, facilitated by the removal of a CTCF-mediated chromatin boundary. This model is consistent with the classic function of CTCF at insulator elements[108,109] and with instances of enhancer hijacking resulting from dissolved chromatin boundaries in ectopic settings[89,110–112] and in cancer[12,88,113]. In addition, CTCF has been shown recently to mediate the minority of tissue-invariant promoter-enhancer contacts that persist independently of enhancer activity[114]. Therefore, genetically-determined CTCF binding, either *in cis* or *in trans* to enhancers, may be a key mechanism for uncoupling enhancer activity and promoter communication. However, the apparent rarity of such direct effects on architectural protein binding reinforces the notion that, at least in the case of communication between active enhancers and promoters, genetic effects on chromosomal contact are more commonly mediated by epigenetic phenomena.

In conclusion, our analysis reveals abundant shared effects of common genetic variants on enhancer activity, connectivity and gene expression in human primary cells, while highlighting the ability of genetic variants to influence gene expression through direct modulation of architectural protein binding. Taken together, our results provide insights into the mechanisms underpinning enhancer function and its genetic perturbation in disease.

Materials and methods

Monocyte eQTL data

To inform our CHi-C design, we used a multi-cohort eQTL analysis of unstimulated monocyte data from a total of 1,480 samples[19]. The samples for this analysis comprised individuals from three cohorts: 1) The Wellcome Trust Centre for Human Genetics (WTCHG), University of Oxford, UK) (N = 432); 2) The CEDAR project, University de Liege, Belgium (N = 300) and 3) The Cardiogenics Consortium eQTL project, University of Leicester and University of Cambridge (N = 758)[18]. The data from this study was subsequently incorporated into the eQTLGen Consortium as a validation of monocyte *trans*-eQTLs[18]. Here, pre-processing, genotype imputation, quality control and eQTL analyses were performed as described[19]. Briefly, significant eQTLs (FDR < 0.01) were detected using a linear mixed model (LIMIX[115] version 0.8.5, PEER version 1.0[116]) and only variants within the gene body or within 1 Mb windows flanking the 5' and 3' of the tested gene were considered. Independent signals were identified using forward selection, which recursively adds the strongest associated variant of the previous association test as a covariate. Up to five iterations of forward selection were performed, thereby identifying up to five independently associated variants with the expression of each eGene[19].

Human DpnII eQTL Capture Hi-C design

We designed a novel eQTL capture Hi-C system enabling us to assay chromatin loops containing enhancer-borne eQTLs at the loop anchors. We found all SNPs in tight LD ("proxy SNPs") with the lead monocyte eQTLs from the analysis described above (R^2 >= 0.9) using PLINK[117], with 1000 Genomes[118] phase 3 (European samples) at MAF 10%. We then located the transcription start sites of the corresponding eGenes (Ensembl Genes release 96) using the Biomart R package and removed all proxy SNPs that were within 10 kb of a TSS for the corresponding eGene. SNPs and TSS were assigned to *DpnII* fragments and we only retained eQTLs and proxy SNPs that were within 100bp of a DpnII cut site, to increase our chances of detecting these SNPs within our CHi-C reads. To enrich our design for SNPs with likely regulatory function, we removed proxy SNPs that did not intersect a known regulatory region (Ensembl Regulatory Build[41] release 96), whilst keeping all remaining lead eQTLs in the design. We also removed 31 regions that had very large LD blocks (>200 kb). Next, focusing on gene promoters, we generated 50 kb "eGene windows'' that were situated from 5 kb downstream of the eGene promoter to 45 kb upstream, with respect to the position of the eQTL LD block. We targeted all gene TSS within these 50 kb windows. Likewise, we generated a 50 kb "mirror window" to the opposite side of the LD block and targeted all TSS within that region (*Fig. 1A*). Thus, we targeted a set of distance-matched control genes for

each eQTL-eGene pairing. We designed 120bp probes to capture the *DpnII* fragments encompassing the SNPs and proxies (one probe per fragment). In total our capture system contained 18,178 probes that targeted 1,458 eQTLs and 2,571 proxy SNPs (a mean of 3 SNPs per eQTL) and 8,533 genes. Furthermore, since we had captured eGene promoters in our study, we were able to include eQTLs and/or proxy SNPs at "other ends" of CHi-C contacts in our analysis. We also directly captured 4,718 distance-matched control genes for these eQTL loci. Note that CHi-C data arising from this array design was further filtered post-hoc prior to contact QTL detection, as described below in "CHi-C data processing for contact QTL detection".

Monocyte purification, crosslinking and lysis

Classical monocytes (CD14+, CD16-) were extracted from apheresis cones discarded after platelet donation by NIHR National BioResource volunteers after obtaining informed consent (REC approval 12/EE/0040). Since sex is known to affect monocyte biology[119–122], we sought to recruit a single-sex cohort. The male sex was chosen based on platelet donor availability at the recruitment centre. Sample-level information is provided in *Table S13*.

Peripheral blood mononuclear cells were isolated by standard Ficoll gradient centrifugation and classical monocytes isolated using EasySep™ Human Monocyte Isolation Kit from Stemcell Technologies according to the manufacturer's instructions. For each sample, 2 million monocytes were frozen in TRIzol and stored at -80°C for later RNA extraction. Meanwhile, between 16.4 and 64.8 million (median 25.3 million) monocytes were crosslinked in 2% formaldehyde for 10 minutes, after which the reaction was quenched with 0.125M glycine, the supernatant removed and the crosslinked cell pellet flash frozen and stored at -80°C. Prior to processing for CHi-C, ATAC-seq and DNA extraction (for genotyping), the crosslinked pellets were thawed on ice and lysed for 30 minutes in Hi-C lysis buffer (10mM Tris-HCl, pH 8.0, 10mM NaCl, 0.2% IGEPAL CA-630 and 1X protease inhibitors) at a density of approximately 3330 cells/µL. The nuclei were then divided into aliquots of 50K - 3 million cells, centrifuged, the supernatant removed and the pellets snap frozen on dry ice and stored at -80°C.

Capture Hi-C

Capture Hi-C libraries (1-2 technical replicates per donor) were generated on 300,000 to 1 million crosslinked nuclei as previously described[123,124]. Following lysis, the nuclei were permeabilised and digested with DpnII (NEB) overnight. The restriction overhangs were filled in using a biotinylated dATP (Jena Bioscience) and ligation was performed for 4 hours at 16°C (T4 DNA ligase; Life Technologies). The crosslinks were reversed using proteinase K and overnight

incubation at 65°C, followed by purification with SPRI beads (AMPure XP; Beckman Coulter). Short fragments up to 1000 bp were produced via tagmentation and the biotinylated restriction junctions were then pulled down using MyOne C1 streptavidin beads (Life Technologies). PCR amplification (5 cycles) was performed on the libraries directly bound to the C-1 beads and the libraries were purified using SPRI beads as before. eQTL Capture was performed using the custom-designed Agilent SureSelect system described above, following the manufacturer's protocol. The libraries were sequenced using 150 bp paired-end sequencing on an Illumina NovaSeq (Novogene UK), NextSeq 500 or HiSeq 2500 (read metrics given in *Table S2*).

CHi-C data processing and detection of significant contacts

The CHi-C reads were processed using a modified version of the Hi-C User Pipeline (HiCUP[125]) v0.7.4, HiCUP Combinations[21] (https://github.com/StevenWingett/HiCUP/tree/combinations), which first creates all possible combinations of ditags from the paired reads prior to mapping and performing standard filtering for common Hi-C artefacts. To determine the correlation between samples, we first filtered our bam files to on-target reads and converted them to .homer files using *hicup2homer* function included with HiCUP, and then created 10 kb interaction matrices using homer[126] v4.10.4. Then, pairwise stratum-adjusted correlation coefficients (SCC) were generated using the R package HiCRep[127] v1.10.0. To generate the deeply sequenced consensus dataset, CHiCAGO input (.chinput) files were generated per replicate, using *bam2chicago.sh* from chicagoTools[20] v1.13. Then, read counts across all .chinput files were summed in R and CHiCAGO was run at *DpnII* resolution and at the level of 5 kb bins, as previously described[20,21]. We performed tests for enrichment at other ends of regions interacting with eQTLs and gene promoters using the *peakEnrichment4Features* function in CHiCAGO using the following datasets: histone modifications in monocytes from ENCODE ChIP-seq data[23] (ENCFF493TKQ, ENCFF933WTT, ENCFF682RJQ, ENCFF570DGJ, ENCFF275IMY, ENCFF473HCC), classical monocyte segmentations downloaded from Ensembl Regulatory Build (http://ftp.ebi.ac.uk/pub/databases/blueprint/releases/current_release/homo_sapiens/secondary_analysis/Ensembl_Regulatory_Build/hg38/projected_segmentations) and CTCF from monocytes in Blueprint[29]. To obtain interactions within individuals, CHiCAGO was also run on each biological replicate separately. To generate locus plots using the CHi-C data and other modalities in this study, we used the PlotGardener[128] R package v1.2.10.

CHi-C ABC

The CHi-C ABC method that we presented recently[24] is a modification of the original Activity-by-Contact (ABC) technique[26] for CHi-C data. Briefly, for a given promoter-enhancer pair, ABC score is defined as a product of Activity (accessibility and epigenetic marks) and Contact frequency (in

the case of CHiC-ABC, based on normalised and imputed CHi-C signals) normalised to the sum of ABC scores for all enhancers within a 5 megabase window surrounding a gene's promoter. Custom scripts for running CHi-C ABC are available at https://github.com/pavarte/PCHIC-ABC-Prediction.

To select a threshold for ABC scores, we drew inspiration from a method by Xu et al.[129] based on optimising Pearson's correlation between gene-level ABC scores and the observed gene expression levels (expressed as CPMs). Specifically, we computed gene-level ABC scores by summing the products of enhancer activity and contact frequencies, including in this calculation only those enhancers that passed a specific "enhancer inclusion" threshold on the standard (enhancer-level) ABC scores. Iterating over a range of enhancer inclusion thresholds, we then assessed Pearson's correlation between gene-level ABC scores and gene expression at each threshold for all genes. As expected, this correlation initially increased with increasing thresholds, as enhancers with baseline levels of activity and/or contact were removed from the calculation, and then started to decrease as enhancers with appreciable impact on gene expression were increasingly filtered out (data not shown). The standard (enhancer-level) ABC score threshold of 0.012 yielded a maximum correlation between gene-level ABC scores and gene expression ($r \sim 0.31$) and was used for downstream analyses. We additionally verified by permuting gene labels that the observed correlation was highly significant (p value < 0.001 over 1000 permutations).

Hi-C and TAD analysis

To generate TADs, Hi-C libraries from two donors underwent 150bp PE sequencing on a NovaSeq (Novogene UK). The reads were processed and filtered using HiCUP v 0.7.4, yielding ~117 and ~127.5 million clean unique reads. TADs were called using *findTADsAndLoops.pl* and the TADs for both replicates were then merged using *merge2Dbed.pl*, both from the homer toolkit[126] v4.10.4. We used the *foverlaps* function in data.table R package to intersect eQTLs (or SNPs in tight LD), along with their target eGene promoters, with TADs. We designated eQTLs and genes "same TAD" if any part of the LD block for the eQTL and the gene promoter intersected the same TAD. Alternatively, we designated them "different TAD" if both the eQTL LD block and the gene promoter intersected different TADs, and no part of the LD block was in the same TAD as the gene promoter. We disregarded loci where the eQTL and/or the gene promoter fell outside of called TADs.

ATAC-seq

ATAC-seq libraries were generated following the omni-ATAC protocol, with modifications for crosslinked chromatin as previously described[130,131]. Between 25 thousand and 100 thousand crosslinked, lysed nuclei (1-2 technical replicates per individual) were re-suspended in a tagmentation reaction mixture (10mM tris HCl [pH 8.4-9.0], 5 mM MgCl2, 10% DMF, 0.005% digitonin and 0.05% tween 20 and up to 5µL of homemade Tn5 enzyme) and incubated at 37°C for 2 hours with 1000 rpm mixing. Reverse crosslinking was performed using proteinase K overnight at 65°C and then samples were cleaned using minElute spin columns (Qiagen). Libraries were amplified using 5-7 cycles of PCR using KAPA HiFi DNA polymerase and purified with 1.1X SPRI (Beckman Coulter). Nucleosome profiles were manually inspected using a Tapestation with the D1000 Screen Tape system (Agilent). Libraries underwent 150 bp PE sequencing on a NovaSeq (Novogene UK) or HiSeq 2500 (read metrics given in *Table S3*).

Adapter contamination was removed from the reads using NGmerge[132] v0.3. Reads were aligned to the GRCh38 genome using Bowtie2[133] v2.2.9 using the settings: *--very-sensitive* and *-X 2000*. Next, mitochondrial reads, multi-mapped reads and non-paired reads were identified and removed using Samtools[134] v1.3.1. For calling peaks, the bam files for technical replicates were merged (where applicable) and PCR duplicates were removed using Picard[135] v2.6.0. Post-filtering we had a minimum of 24.9 million and a median of 34.7 million unique reads per individual. Correlation between biological replicates was determined using the *plotCorrelation* function from DeepTools[136] v3.3.1 (*Fig. S1B*). For each sample, we confirmed the presence of distinct fragment sizes corresponding to nucleosomes and confirmed the enrichment of nucleosome-free peaks at TSSs and mononucleosome peaks surrounding TSSs using the R package soGGi[137] v1.18.0 (representative examples shown in *Fig. S1C*). We combined all sample BAM files using Samtools *merge* to create a consensus dataset and ran the hidden Markov Model HMMRATAC[22] v1.2.10 to detect peaks of open chromatin. In order to produce signal tracks for visualisation, the merged ATAC-seq samples were additionally processed by the Genrich peak caller[138] v0.5_dev, which produces bedGraph files of p values and read pileups.

RNA-seq

RNA was extracted from TRIzol samples (2 million monocyte cells) using chloroform in MaXtract High Density 2mL tubes, followed by dissociation with QIAShredder columns and purification using the RNeasy kit with RNase-free DNase treatment (all Qiagen). RNA samples were quantified with a Qubit RNA assay (Thermo Fisher Scientific) and 800ng RNA was submitted to Novogene UK for mRNA-seq library preparation (poly A enrichment) and 150 bp PE sequencing obtaining

~20 million reads per sample (*Table S4*). Read QC and mapping were performed according to the Novogene bioinformatics pipeline. Reads were mapped to the GRCh38 genome using STAR[139] v2.6.1d allowing for 2 mismatches and assigned to Ensembl Release 94 genes. Pearson's correlation of gene FPKM values was determined between all pairwise samples (*Figure S1D*).

Determining the relationship between gene expression and enhancer-promoter contacts

To probe the relationship between gene expression and enhancer-promoter contacts, we first calculated transcripts per million (TPM) for every gene in each individual. We determined significant CHi-C interactions between gene promoters and open chromatin by intersecting the peak matrix produced by CHiCAGO, which contained significant interactions at 5kb binned resolution in each individual (CHiCAGO score >= 5), with the locations of captured gene promoters on one end and open chromatin on the other end (called by HMMRATAC in the consensus ATAC-seq dataset). We summed the number of active interacting regions per gene promoter in each individual and determined Spearman's rank correlation between log(no. of interacting regions) and log(gene TPM).

Genotyping

DNA was extracted from 2 million crosslinked, lysed nuclei using the QiaAmp FFPE kit (Qiagen), starting from the Proteinase K step. Genotyping was performed on the Infinium Global Screening Array (Illumina) by The IoPPN Genomics & Biomarker Core Facility at King's College, London. Initial genotyping QC was performed using Zcall and all samples passed the minimum call rate of 95%. Prior to imputation, variants were filtered to 5% minor allele frequency (MAF) and 100% call rate in our cohort. We removed SNPs deviating from Hardy-Weinberg equilibrium at a threshold of 1E-05. The remaining 447,026 variants underwent genome-wide imputation on the Michigan Imputation Server[140] v1.2.4 using the Haplotype Reference Consortium[141] with the following programs: minimac4[142] v1.0.2 and phasing with Eagle[143] v2.4. Post-imputation, the variants were filtered for R^2 >= 0.3 and 5% MAF, resulting in a total of 5,380,837 SNPs in our cohort. We retained the information on imputation probabilities and incorporated them into the BaseQTL analysis.

Validation of eQTL effects

To ascertain the concordance between the eQTL effects in the multi-cohort eQTL study and our monocyte cohort, we detected eQTLs in our RNA-seq data using two different approaches: 1) Matrix eQTL[144] and 2) BaseQTL[28]. For Matrix eQTL, total gene counts were determined using FeatureCounts[145] v1.5.0-p3 with default settings in the Novogene analysis pipeline. The total read counts were then normalised using *rlog* in the DESeq2[146] R package v1.26.0 and Matrix eQTL[144] was run using sample genotypes. For BaseQTL, raw RNA-seq reads were processed using the full BaseQTL input Snakemake pipeline for RNA-seq data (available from https://gitlab.com/evigorito/baseqtl_pipeline/-/tree/master/input), which performs mapping using STAR[139], generates allele-specific counts using phASER[147] and determines the baseline allelic imbalance (AI) of each variant, which is then incorporated in the BaseQTL analysis. BaseQTL exploits allele specific expression (ASE) to increase power. However, homozygous individuals erroneously genotyped as heterozygous using DNA would overestimate the allelic imbalance observed from RNA-seq reads, which may lead to false positive calls. To mitigate this source of error we employed two strategies. First, we genotyped exonic SNPs from RNA-seq reds using BBmix[148], so genotypes and allelic imbalance were estimated from the same source. Second, BaseQTL for calling eQTLs has been designed to use hard genotype calls. To minimise the risk of false positive calls due to genotype miscalls we only included candidate eQTLs with a genotype probability equal or greater than 0.99. BaseQTL was then run using default parameters except that we allowed for the minimum number of heterozygous individuals to be 3 (defaults 5) in order to run the model, and the minimum number of heterozygous individual with ASE information to be 3 (defaults to 5) in order to run to model ASE. Finally, we intersected MatrixQTL or BaseQTL results with lead eQTLs from the multi-cohort study and determined Spearman's correlation of the effect sizes (*Figure S2*).

CHi-C data processing for contact QTL detection

We applied post-hoc processing of our CHi-C data for contact QTL detection. First, we discarded all eQTLs that had a proxy SNP in tight LD that was proximal to the eGene promoter (< 10 kb), in order to assay only truly distal effects, accounting for LD. In addition, we filtered out all eQTLs with potential effects on *DpnII* cut sites (either in the reference or alternative allele), that could cause allelic imbalance due to either 1) the technical ability of *DpnII* to fragment the library or 2) read mapping bias. In this filtering step, we also accounted for SNPs in LD that intersected the same DpnII fragment and combinations of SNPs within 4bp that could potentially create a cut site (*GATC*). Finally, we estimated the affinity of our SureSelect capture probes for the *DpnII* CHi-C fragments containing either the reference or alternative haplotypes (all variants in R^2 >= 0.9 with

the lead eQTL) as the melting temperatures of the corresponding hybrids using the Tm_GC function from the R package TmCalculator v1.0.1, assuming Tris=20 and Mg=308. Haplotype-specific differences in the affinity were 0.18% on average and did not exceed 0.73% in all cases (data not shown), and therefore we assumed them to have a negligible effect on capture efficiency. After filtering and processing, our total possible eQTL discovery set encompassed 1,197 eQTLs and 12,295 proxy SNPs, either by direct capture or through capturing the eGene promoter (N = 1,074 corresponding eGenes).

BaseQTL detection of contact QTLs and ATAC QTLs

Prior to running BaseQTL on the CHi-C and ATAC-seq modalities, we first ran the refBias section of the BaseQTL input workflow, using the scripts available from Gitlab (https://gitlab.com/evigorito/baseqtl_pipeline/-/tree/master/input/refbias). For each modality we used the de-duplicated bam files and all genotyped variants passing the quality thresholds described above.

We processed the CHi-C reads such that we could detect 1) genetic effects on eQTL-eGene contacts and 2) genetic effects on non-eQTL-Gene contacts. For eQTL-eGene contacts, we processed the de-duplicated and filtered CHi-C bam files such that we only retained reads involving the *DpnII* fragment containing an eQTL (or a proxy SNP in tight LD, R^2 >= 0.9) and a 5 kb region encompassing the eGene promoter. Where eGenes had more than one promoter, we compiled all reads across promoters. We did not require evidence of significant contact above background. For non-eQTL variants, we looked for variants within other-end *DpnII* fragments of interactions involving captured gene promoters in our analysis (CHiCAGO score >= 5 in the consensus dataset). These variants were not associated with expression in the multi-cohort eQTL analysis[18,19], nor did they form significant interactions with eQTLs captured in our study design, thus there was no evidence for a role in regulating gene expression. To filter the bam files we first converted them to bedpe using the *bamtobed* function from BEDtools[149] v2.30.0. Then, we filtered the resultant bedpe against a reference bedpe with the paired regions that we wanted to test, using the *pairtopair* function from BEDtools. With the resultant read IDs we were then able to filter the original CHi-C bam files using SAMtools[149,150].

We extended BaseQTL to detect QTLs from CHi-C and ATAC reads. For these modalities, we model the uncertainty on the genotype calls obtained by imputation. BaseQTL combines total reads mapping interactions across individuals modelled by a negative binomial distribution f_{NB} with reads mapping the alternative allele in heterozygous individuals modelled by a beta binomial distribution f_{BB}. The model estimates the logarithm of the allelic fold change. We define the

allelic fold change β_{aFC} as the ratio of the expected number of contacts in individuals homozygous for the alternative allele divided by the expected number of contacts in individuals homozygous for the reference allele. In contrast to the original development case for BaseQTL[28], where we had to accommodate inferred haplotypes at the feature (expression across the whole gene body) as well as the regulatory SNP, which caused us to restrict to using genotypes called with high certainty, here the feature (contact or chromatin accessibility) is measured simply by read depth, which allowed us to make wider use of uncertain genotypes called through imputation. Using the same notation as Vigorito et al.[28], we adapted the likelihood as follows:

$$L(\beta_{aFC}, \gamma, \theta, \phi | c, n_1, G, X) \propto \prod_{i=1}^{N} [p(G_i = 1) f_{NB}(c_i | \beta_{aFC}, \phi, X) f_{BB}(n_{1i}; \pi, \theta, c_i) + \sum_{G_i=0,2} p(G_i = g) f_{NB}(c_i | \beta_{aFC}, \phi, X)];$$

$$\pi = \frac{exp(\alpha_{0i} + \beta_{aFC})}{1 + exp(\alpha_{0i} + \beta_{aFC})} \; for\ G_i\ heterozygous,$$

$$\pi = \frac{exp(\alpha_{0i})}{1 + exp(\alpha_{0i})} \; for\ G_i\ homozygous.$$

Where:
γ is a vector of the regression parameters for covariates
ϕ is the overdispersion parameter for the negative binomial distribution
θ is the overdispersion parameter for the beta binomial distribution
c_i is the total counts mapping the feature for individual i
n_{1i} is the number of counts mapping the alternative allele of the candidate QTL in individual i
X is a matrix of covariates
G_i is the genotype of the candidate QTL for individual i
$p(G_i = g)$ is the genotype probability for each genotype class obtained from imputation
α_{0i} is used to adjust for reference mapping bias, would be 0 in the absence of bias (Baseqtl ref)

We used the cis-eQTL GTEx-derived default prior[151] on β_{aFC} in BaseQTL, which is a mixture of two Gaussian distributions (N(0, 0.001) and N(0, 0.121)), except that for this application we considered a 50/50 mixture to reflect our expectation that 50% of the candidate QTLs have true effects.

To obtain the number of reads mapping the alternative allele of the candidate QTL in heterozygous individuals we used phASER[147] v1.1.1 from the filtered CHi-C bam files and ATAC-seq bam files, respectively. The same bam files were used to aggregate reads overlapping each feature. The code with the model for running BaseQTL on these modalities can be found at https://gitlab.com/evigorito/baseqtl_atac_chic_eqtl. For all the modalities analysed by BaseQTL,

significant associations were considered as those for which zero was excluded from the 99% credible interval for β_{aFC}. In addition, we discarded variants identified via allelic imbalance in heterozygotes that had evidence of substantial baseline bias towards the reference genome (AI estimate < 0.4) and variants for which the model failed to converge (Rhat >= 1.01).

Comparison of effects on promoter contact between eQTLs and non-eQTLs

We compared the BaseQTL results for eQTL-eGene contacts with non-eQTL-gene contacts in the following manner. First, we ensured that the non-eQTL variants (or their proxies in tight LD, R^2 > 0.9) were not associated with differential *DpnII* activity, as described in the CHi-C design above. Then, to rank candidate QTLs we calculated the probability that the candidate SNP is a true eQTL. If we denote with $r_i \in \{0,1\}$ the unknown truth for a SNP being (1) or not (0) a QTL, we can calculate $p_i = P(r_i = 1|data)$ from our posterior samples by calculating the proportion of times that 0 was excluded from the 99% credible interval. To do this with a manageable number of samples, we used a normal approximation to the marginal posterior distribution of the QTL effect. We denote this quantity as approx.pp, and we use it to rank SNPs. We computed $P(r_i = 0|data)$ = 1- $P(r_i = 1|data)$ and then used a permutation test to assess whether eQTL-eGene contacts have a lower mean approx.pp compared with non-eQTL-gene contacts. The null distribution of mean approx.pp was obtained by generating 10,000 samples of 800 non-eQTL gene contacts each and taking the mean of the corresponding approx.pp for each sample. To control for read coverage that confers different statistical power of association detection, we sampled only those non-eQTL-gene contacts whose read counts mapped to the same quantile of the read count distribution of the corresponding eQTL-eGene contact.

Detection of trimodal QTLs by the GUESS framework

To generate the regions for GUESS, windows were defined at regions in close proximity to eQTL-containing enhancers targeted in the CHi-C design and peaks of open chromatin in the ATAC-seq data. Specifically, we expanded 5 kb on either side from the CHi-C baited *DpnII* fragments and from the ATAC-seq peaks and took the minimum intersection of these two ranges. We considered all genotyped variants following within these windows for the GUESS analysis. To avoid the impact of *DpnII* cut site perturbation in the CHi-C data, we removed windows where any of the considered variants affected or generated a DpnII cut site at the CHi-C bait fragment. We additionally did not consider windows where any of the variants fell within 10 kb of a promoter for the eGene.

We performed rlog transformation on the total counts for CHi-C, ATAC and RNA-seq using DESeq2[146]. For each window, we then ran GUESS[30,31] for 20,000 iterations, of which 5,000 as burn-in with 3 parallel chains to avoid the algorithm being stuck in a local mode. A priori, and independently of the number of genetic variants included in each window, the expected number of trimodal QTL was set equal to 1 as well as its standard deviation which implies a range between zero and four for the number of significant trimodal QTLs. Prior to the analysis, the genetic variants were standardised to place appropriate prior mass on reasonable values of the non-zero regression coefficients. Moreover, we subtracted the mean of the rlog transformed total counts, implicitly specifying a non-informative prior on the intercept for each molecular trait. We did not filter genetic variants by LD, since the GUESS algorithm is robust to extreme multicollinearity given that marginal likelihood is calculated using the technique of QR matrix decomposition and parallel tempering allows the inclusion of collinear predictors in different chains but not on the same one. Swapping selected genetic variants between chains allows the algorithm to explore the space of collinear predictors without incurring rank deficiency. For each window analysed, and the chain with temperature equal to one, we recorded the best model visited, i.e., the combination of genetic variants visited by the algorithm with the largest marginal likelihood score, and the marginal posterior probability of inclusion, i.e., the marginal association strength of each genetic variant with all molecular traits. If the best model visited was the null model, we declared no associations. Otherwise, we calculated the FDR threshold for the marginal posterior probability of inclusion as

$$arg \max_{j} \sum_{j} (1 - mPPI_{(j)}) \leq q,$$

where $mPPI_{(j)}$ is the jth-ordered marginal posterior probability of inclusion sorted in non-decreasing order and q is the designed level of False Discovery Rate, set at 0.05[32]. While the best model visited can overlap with the model obtained by filtering the marginal posterior probability of inclusion by the specified FDR level, in general, they do not coincide. We also generated the posterior effects size and its standard error averaging over all models visited, in a post-processing step, thus performing Bayesian Model Averaging[152]. Finally, in analogy with classical statistics, we defined the posterior expectation of the z-score as

$$\mathrm{E}(z_{jk}|Y) \approx \frac{\mathrm{E}(\beta_{jk}|Y)}{\{\mathrm{E}(\sigma^2(\beta_{jk})|Y)\}^{1/2}},$$

where β_{jk} is the effect size of the jth genetic variant on the kth molecular trait in each window and $\sigma^2(\beta_{jk})$ is the variance of the effect size for the jth genetic variant on the kth molecular trait.

We further filtered the GUESS results after FDR selection as follows: firstly, we discarded variants that had evidence of baseline allelic imbalance at the feature in question (AI estimate < 0.4), as determined by the refBias pipeline of BaseQTL. Finally, we filtered out QTLs with very low absolute effect sizes to ensure the replicability of the results with BaseQTL. For this, we examined the external concordance of variant effects on CHi-C, ATAC and RNA reads between GUESS and BaseQTL (illustrated in *Fig. S4*). We tuned the cutoff of the beta value to maximise this concordance without losing sensitivity, resulting in the following beta cutoffs on significant trimodal QTLs: CHiC beta >= 0.06, ATAC beta >= 0.001 and RNA beta >= 0.004.

To validate a subset of GUESS results with a different statistical strategy, we performed multivariate analysis of variance (MANOVA)[68] on the lead trimodal QTL identified by GUESS in each window, treating the other trimodal QTLs in the same window (if detected at 5% FDR) as covariates (MANCOVA). We also assessed how many of the lead trimodal QTLs could be validated using a univariate approach (ANOVA/ANCOVA) (*Fig. S4B-C*).

Downstream analysis of contact QTLs

For the following analyses, we combined the contact eQTLs from BaseQTL and the trimodal QTLs from GUESS, making a total of 627 QTLs.

Analysis of public ChIP-seq data

Non-redundant peaks for all cell types and factors in the Remap 2022 catalogue[33] were downloaded in GRCh38 from https://remap.univ-amu.fr/download_page. The QTLs were run in ReMapEnrich R package v0.99.0 using a Test Universe consisting of all the locations of all variants submitted to the GUESS and BaseQTL analyses in GRCh38 (filtered for *DpnII* effects and distal from eGene promoters), using 100 shuffles.

Prediction of TF binding perturbation by cQTLs

The prediction of transcription factor (TF) binding was executed using the Enformer model[153]. We curated the 5313 genomic tracks in the Enformer training set to focus on 1433 ChIP-seq tracks for 662 TFs, excluding tracks for chromatin co-factors and histone modifying enzymes (*Table S3*). Given the scarcity of TF ChIP-Seq data generated in monocytes in the Enformer training set, we used data across all cell types and conditions.

Since our QTL calling procedures accounted for effects of LD, we chose to test the effects of individual cQTLs rather than of the haplotypes linked with them at a predefined LD threshold, although our chosen approach is likely more conservative. For each variant of interest, we supplied Enformer with two copies of a 393,216 bp sequence centred at each variant of interest incorporating the reference or alternative allele of the variant, respectively. Based on these input sequences, Enformer yielded TF binding scores for each of the 128 bp bins within the central 114,688 bps. The scores for the two central bins (i.e., those flanking the variant of interest) were averaged to obtain the raw scores for each variant, which will be referred to as the "ref allele score" and "alt allele score". These scores were then used to calculate TF binding perturbation scores in a manner that considers both the absolute and relative allelic changes in the binding prediction score[154]:

$$\text{TF binding perturbation score} = \text{SAD}*\text{SAR},$$

where

$$\text{SAD (SNP Activity Difference)} = \text{alt allele score} - \text{ref allele score},$$

$$\text{SAR (SNP Activity Ratio)} = \log_2(\text{alt allele score}+0.001) - \log_2(\text{ref allele score}+0.001)$$

To define a threshold on the TF binding perturbation scores in a consistent manner for each TF and cell type, we obtained the distribution of these scores in each genomic track for variants randomly sampled from the dbSNP database and used it to define empirical p-values for the variants of interest. Variants with FDR-corrected p values (Benjamini-Hochberg) below 0.05 in at least one data track for a specific factor (in any cell type and condition) were considered as significantly perturbing the binding of the respective TF.

To assess jointly perturbed TF binding events, we used hierarchical clustering on the binary matrix of variant effects on each TF (generated as described above) using the Gower distance metric implemented in the R function *daisy* and the default clustering method ("complete") in *hclust*. The resulting tree was partitioned into 20 clusters using the *cutree* function in R.

Intersection of cQTLs with ATAC-seq TF footprints, ChIP signals and binding motifs

ATAC-seq footprints were detected using the TOBIAS framework[155]. We ran the TOBIAS Snakemake pipeline (available at https://github.com/loosolab/TOBIAS_snakemake) on clean bam files for all 34 ATAC-seq replicates and detected the binding of all TFs with motifs in the Jaspar 2020 dataset[156] at ATAC-seq footprints. For each TF we took the bed files for "bound" peaks and intersected them with the 256 bp windows surrounding each cQTL (i.e. 128 bp

upstream and downstream). The enrichment for each TF was determined using the Remapenrich R package v0.99.0 using 100 shuffles within the same chromosome as each cQTL.

For cQTLs predicted to perturb the binding of CTCF, GABPA or SPI1 binding, we additionally calculated TF binding motif matches in the same 256bp windows based on the position weight matrices from the HOCOMOCOv11 database[157] using the R package motifmatchr at a score threshold of 1e-4. We also assessed the ChIP signals for the respective TFs in the same 256-bp windows using data from the ReMap 2022 catalogue[33]. The enrichments of these cQTLs for the motifs and/or ChIP peaks of CTCF, GABPA or SPI1, respectively, were calculated as follows:

$$E(P/NP)_{TF} = (\#cQTLs_{P(TF),motif|ChIP}/\#cQTLs_{P(TF)})/(\#cQTLs_{NP(TF),motif|ChIP}/\#cQTLs_{NP(TF)}),$$

where *P(TF)* and *NP(TF)* refer to cQTLs predicted and not predicted to perturb the binding of a specific TF, respectively, and *motif|ChIP* refers to cQTLs with a motif match and/or ChIP signal for the same TF.

Disruption of TF binding motifs by cQTLs

TF binding motif disruption by cQTLs was detected using the motifbreakR package[158] in R using the motif data sources “HOCOMOCOv11-core-A”, “HOCOMOCOv11-core-B”, “HOCOMOCOv11-core-C” from the R package MotifDb. Motif matches were detected at a p-value threshold of 1e-4 using the “ic” method against a flat genomic background. The predicted “strong” variant effects on the TF binding motif (corresponding to the min absolute difference in position weight matrix match scores of 0.7) were used in downstream analysis.

Comparison of predicted TF binding perturbation patterns at contact versus ATAC QTLs

We defined a negative set of 513 “non-contact” ATAC eQTLs by selecting the eQTL variants from the multi-cohort analysis[19] that were detected as ATAC eQTLs (BaseQTL 99% CI), but not as contact eQTLs (BaseQTL 95% CI) or trimodal QTLs (GUESS 5% FDR). TF binding perturbation predictions were obtained for this negative set using Enformer in the same manner as for cQTLs (see above). We then used logistic LASSO implemented in R package glmnet[159] on the merged dataset using the binary effect of a QTL on contact (1: cQTL; 0: ATAC eQTL) as the outcome, and dummy variables representing the 662 binarised TF binding perturbation predictions as covariates. The shrinkage parameter lambda was selected by cross-validation using the function *cv.glmnet* from the glmnet package (lambda.min was used). Random forest classification was performed using the randomForest R package. AUCs were generated using the R package pROC[160].

GWAS intersection

For each cQTL, we identified proxy variants at R^2 >= 0.8 using PLINK[117] with a 1000 Genomes reference panel[118]. We then obtained a direct overlap of cQTLs and the proxy variants with variants listed in the GWAS Catalog[161] (downloaded on 16th June 2023).

Data and code availability

The code used to generate the results in this study are available at the following repositories: https://github.com/FunctionalGeneControl/contactQTLs (eQTL CHi-C design and data processing)), https://github.com/pavarte/PCHIC-ABC-Prediction (CHi-C ABC analysis), https://gitlab.com/evigorito/baseqtl_atac_chic_eqtl (BaseQTL extended to CHi-C and ATAC-seq modalities) and https://github.com/lb664/R2GUESS (GUESS framework).

The raw data generated in this study will be submitted to the European Genome-Phenome Archive (EGA) under managed access, as outlined in the donor terms of consent. Large supplementary data (eQTL CHi-C design, CHi-C significant interactions, ATAC-seq peaks, TF footprinting, RNA-seq counts) are available on the Open Science Framework (OSF) at https://osf.io/szntj/

Acknowledgements

The authors would like to thank Oliver Stegle, Laurence Game and Ivan Andrew for advice and technical assistance, and Matthias Merkenschlager for the critical reading of the manuscript.

We thank UK's National Institute for Health and Care Research (NIHR) BioResource volunteers for their participation, and gratefully acknowledge NIHR BioResource centres, UK's National Health Service (NHS) Trusts and staff for their contribution. We thank the NIHR, NHS Blood and Transplant, and Health Data Research UK as part of the Digital Innovation Hub Programme. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

This work was supported by the Medical Research Council (MRC) of the UK through MRC Investigator funding to M.S. (MC-A652-5QA20). M.F. is supported by the British Heart Foundation (BHF; FS/18/53/33863), the BHF Cambridge Centre for Research Excellence (RE/18/1/34212) and the NIHR Exeter Biomedical Research Centre. C.W and E.V. are supported by the MRC

(MC_UU_00002/4) and C.W. by the Wellcome Trust (WT220788). L.B. gratefully acknowledges funding from the MRC (MR/W029790/1), the Alan Turing Institute under the Engineering and Physical Sciences Research Council (EPSRC) grant EP/N510129/1 and the Marmaduke Sheild Fund.

Conflict of Interest

CW is supported by GSK and MSD and is a part time employee of GSK. GSK had no role in this study or the decision to publish. MS is a shareholder of Enhanc3D Genomics Ltd.

References

1. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet*. 2014;15(4):272-286.

2. Grosveld F, van Staalduinen J, Stadhouders R. Transcriptional Regulation by (Super)Enhancers: From Discovery to Mechanisms. *Annu Rev Genomics Hum Genet*. 2021;22:127-146.

3. Ray-Jones H, Spivakov M. Transcriptional enhancers and their communication with gene promoters. *Cell Mol Life Sci*. 2021;78(19-20):6453-6485.

4. Schoenfelder S, Fraser P. Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet*. 2019;20(8):437-455.

5. Wurmser A, Basu S. Enhancer-Promoter Communication: It's Not Just About Contact. *Front Mol Biosci*. 2022;9:867303.

6. Lim B, Levine MS. Enhancer-promoter communication: hubs or loops? *Curr Opin Genet Dev*. 2021;67:5-9.

7. Jin F, Li Y, Dixon JR, et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature*. 2013;503(7475):290-294.

8. Javierre BM, Burren OS, Wilder SP, et al. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell*. 2016;167(5):1369-1384.e19.

9. Freire-Pritchett P, Schoenfelder S, Várnai C, et al. Global reorganisation of cis-regulatory units upon lineage commitment of human embryonic stem cells. Published online March

23, 2017. doi:10.7554/eLife.21926

10. de Laat W, Duboule D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature*. 2013;502(7472):499-506.

11. Maurano MT, Humbert R, Rynes E, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337(6099):1190-1195.

12. Claringbould A, Zaugg JB. Enhancers in disease: molecular basis and emerging treatment strategies. *Trends Mol Med*. 2021;27(11):1060-1073.

13. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020;369(6509):1318-1330.

14. Tewhey R, Kotliar D, Park DS, et al. Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay. *Cell*. 2016;165(6):1519-1529.

15. Chandra V, Bhattacharyya S, Schmiedel BJ, et al. Promoter-interacting expression quantitative trait loci are enriched for functional genetic variants. *Nat Genet*. 2021;53(1):110-119.

16. Jung I, Schmitt A, Diao Y, et al. A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nat Genet*. 2019;51(10):1442-1449.

17. Heinz S, Romanoski CE, Benner C, et al. Effect of natural genetic variation on enhancer selection and function. *Nature*. 2013;503(7477):487-492.

18. Võsa U, Claringbould A, Westra HJ, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet*. 2021;53(9):1300-1310.

19. Kreuzhuber R. *The Effect of Non-Coding Variants on Gene Transcription in Human Blood Cell Types*. Doctor of Philosophy (PhD). University of Cambridge; 2019. Accessed July 3, 2023. https://www.repository.cam.ac.uk/handle/1810/292016

20. Cairns J, Freire-Pritchett P, Wingett SW, et al. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol*. 2016;17(1):127.

21. Freire-Pritchett P, Ray-Jones H, Della Rosa M, et al. Detecting chromosomal interactions in Capture Hi-C data with CHiCAGO and companion tools. *Nat Protoc*. 2021;16(9):4144-4176.

22. Tarbell ED, Liu T. HMMRATAC: a Hidden Markov ModeleR for ATAC-seq. *Nucleic Acids Res*. 2019;47(16):e91-e91.

23. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human

genome. *Nature*. 2012;489(7414):57-74.

24. Malysheva V, Ray-Jones H, Cazares TA, et al. High-resolution promoter interaction analysis in Type 3 Innate Lymphoid Cells implicates Batten Disease gene*CLN3*in Crohn's Disease aetiology. *bioRxiv*. Published online October 21, 2022. doi:10.1101/2022.10.19.512842

25. Song M, Yang X, Ren X, et al. Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat Genet*. 2019;51(8):1252-1262.

26. Fulco CP, Nasser J, Jones TR, et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat Genet*. 2019;51(12):1664-1669.

27. Nasser J, Bergman DT, Fulco CP, et al. Genome-wide enhancer maps link risk variants to disease genes. *Nature*. 2021;593(7858):238-243.

28. Vigorito E, Lin WY, Starr C, Kirk PDW, White SR, Wallace C. Detection of quantitative trait loci from RNA-seq data with or without genotypes using BaseQTL. *Nat Comput Sci*. 2021;1:421-432.

29. Chen L, Ge B, Casale FP, et al. Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell*. 2016;167(5):1398-1414.e24.

30. Bottolo L, Chadeau-Hyam M, Hastie DI, et al. GUESS-ing polygenic associations with multiple phenotypes using a GPU-based evolutionary stochastic search algorithm. *PLoS Genet*. 2013;9(8):e1003657.

31. Liquet B, Bottolo L, Campanella G, Richardson S, Chadeau-Hyam M. R2GUESS: A Graphics Processing Unit-Based R Package for Bayesian Variable Selection Regression of Multivariate Responses. *J Stat Softw*. 2016;69(2). doi:10.18637/jss.v069.i02

32. Müller P, Parmigiani G, Rice K. FDR and Bayesian Multiple Comparisons Rules. In: *Proc. Valencia / ISBA 8th World Meeting on Bayesian Statistics*. ; 2006.

33. Hammal F, de Langen P, Bergon A, Lopez F, Ballester B. ReMap 2022: a database of Human, Mouse, Drosophila and Arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments. *Nucleic Acids Res*. 2021;50(D1):D316-D325.

34. Consiglio CR, Gollnick SO. Androgen Receptor Signaling Positively Regulates Monocytic Development. *Front Immunol*. 2020;11:519383.

35. Rao S, Ahmad K, Ramachandran S. Cooperative binding between distant transcription factors is a hallmark of active enhancers. *Mol Cell*. 2021;81(8):1651-1665.e4.

36. Avsec Ž, Weilert M, Shrikumar A, et al. Base-resolution models of transcription-factor

binding reveal soft motif syntax. *Nat Genet*. 2021;53(3):354-366.

37. Spitz F, Furlong EEM. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet*. 2012;13(9):613-626.

38. Momozawa Y, Dmitrieva J, Théâtre E, et al. IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nat Commun*. 2018;9(1):2427.

39. Jansen R, Hottenga JJ, Nivard MG, et al. Conditional eQTL analysis reveals allelic heterogeneity of gene expression. *Hum Mol Genet*. 2017;26(8):1444-1451.

40. Lepik K, Annilo T, Kukuškina V, et al. C-reactive protein upregulates the whole blood expression of CD59 - an integrative analysis. *PLoS Comput Biol*. 2017;13(9):e1005766.

41. Zerbino DR, Wilder SP, Johnson N, Juettemann T, Flicek PR. The ensembl regulatory build. *Genome Biol*. 2015;16:56.

42. Lonsdale J, Thomas J, Salvatore M, et al. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580-585.

43. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2018;47(D1):D1005-D1012.

44. Geiger MA, Guillaumon AT, Paneni F, Matter CM, Stein S. Role of the Nuclear Receptor Corepressor 1 (NCOR1) in Atherosclerosis and Associated Immunometabolic Diseases. *Front Immunol*. 2020;11:569358.

45. Kannan K, Stewart RM, Bounds W, et al. Lysosome-associated membrane proteins h-LAMP1 (CD107a) and h-LAMP2 (CD107b) are activation-dependent cell surface glycoproteins in human peripheral blood mononuclear cells which mediate cell adhesion to vascular endothelium. *Cell Immunol*. 1996;171(1):10-19.

46. Sproston NR, Ashworth JJ. Role of C-Reactive Protein at Sites of Inflammation and Infection. *Front Immunol*. 2018;9:754.

47. Tansey MG, Wallings RL, Houser MC, Herrick MK, Keating CE, Joers V. Inflammation and immune dysfunction in Parkinson disease. *Nat Rev Immunol*. 2022;22(11):657-673.

48. Ishihara K, Hirano T. BST-1/CD157 regulates the humoral immune responses in vivo. *Chem Immunol*. 2000;75:235-255.

49. Audrito V, Messana VG, Deaglio S. NAMPT and NAPRT: Two Metabolic Enzymes With Key Roles in Inflammation. *Front Oncol*. 2020;10:358.

50. Vuckovic D, Bao EL, Akbari P, et al. The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell*. 2020;182(5):1214-1231.e11.

51. Satake W, Nakabayashi Y, Mizuta I, et al. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat Genet*. 2009;41(12):1303-1307.

52. Nalls MA, Pankratz N, Lill CM, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet*. 2014;46(9):989-993.

53. International Parkinson Disease Genomics Consortium, Nalls MA, Plagnol V, et al. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet*. 2011;377(9766):641-649.

54. Saad M, Lesage S, Saint-Pierre A, et al. Genome-wide association study confirms BST1 and suggests a locus on 12q24 as the risk loci for Parkinson's disease in the European population. *Hum Mol Genet*. 2011;20(3):615-627.

55. Chang D, Nalls MA, Hallgrímsdóttir IB, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet*. 2017;49(10):1511-1516.

56. Pankratz N, Beecham GW, DeStefano AL, et al. Meta-analysis of Parkinson's disease: identification of a novel locus, RIT2. *Ann Neurol*. 2012;71(3):370-384.

57. Grover S, Kumar Sreelatha AA, Pihlstrom L, et al. Genome-wide Association and Meta-analysis of Age at Onset in Parkinson Disease: Evidence From the COURAGE-PD Consortium. *Neurology*. 2022;99(7):e698-e710.

58. Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol*. 2019;18(12):1091-1102.

59. Foo JN, Chew EGY, Chung SJ, et al. Identification of Risk Loci for Parkinson Disease in Asians and Comparison of Risk Between Asians and Europeans: A Genome-Wide Association Study. *JAMA Neurol*. 2020;77(6):746-754.

60. Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet*. 2016;48(7):709-717.

61. Sakaue S, Kanai M, Tanigawa Y, et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet*. 2021;53(10):1415-1424.

62. Chen MH, Raffield LM, Mousas A, et al. Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. *Cell*. 2020;182(5):1198-

1213.e14.

63. Kachuri L, Jeon S, DeWan AT, et al. Genetic determinants of blood-cell traits influence susceptibility to childhood acute lymphoblastic leukemia. *Am J Hum Genet*. 2021;108(10):1823-1835.

64. Astle WJ, Elding H, Jiang T, et al. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell*. 2016;167(5):1415-1429.e19.

65. Greenwald WW, Chiou J, Yan J, et al. Pancreatic islet chromatin accessibility and conformation reveals distal enhancer networks of type 2 diabetes risk. *Nat Commun*. 2019;10(1):2078.

66. Gorkin DU, Qiu Y, Hu M, et al. Common DNA sequence variation influences 3-dimensional conformation of the human genome. *Genome Biol*. 2019;20(1):255.

67. Watt S, Vasquez L, Walter K, et al. Genetic perturbation of PU.1 binding and chromatin looping at neutrophil enhancers associates with autoimmune disease. *Nat Commun*. 2021;12(1):2298.

68. Hand DJ, Taylor CC. *Multivariate Analysis of Variance and Repeated Measures*. Chapman & Hall/CRC; 1987.

69. Spivakov M, Akhtar J, Kheradpour P, et al. Analysis of variation at transcription factor binding sites in Drosophila and humans. *Genome Biol*. 2012;13(9):R49.

70. Lagha M, Bothma JP, Levine M. Mechanisms of transcriptional precision in animal development. *Trends Genet*. 2012;28(8):409-416.

71. Kvon EZ, Waymack R, Gad M, Wunderlich Z. Enhancer redundancy in development and disease. *Nat Rev Genet*. 2021;22(5):324-336.

72. Cannavò E, Khoueiry P, Garfield DA, et al. Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Curr Biol*. 2016;26(1):38-51.

73. Wang X, Goldstein DB. Enhancer Domains Predict Gene Pathogenicity and Inform Gene Discovery in Complex Disease. *Am J Hum Genet*. 2020;106(2):215-233.

74. Spivakov M. Spurious transcription factor binding: non-functional or genetically redundant? *Bioessays*. 2014;36(8):798-806.

75. Dehingia B, Milewska M, Janowski M, Pękowska A. CTCF shapes chromatin structure and gene expression in health and disease. *EMBO Rep*. 2022;23(9):e55146.

76. Akıncılar SC, Khattar E, Boon PLS, Unal B, Fullwood MJ, Tergaonkar V. Long-Range

Chromatin Interactions Drive Mutant TERT Promoter Activation. *Cancer Discov*. 2016;6(11):1276-1291.

77. Ha SD, Cho W, DeKoter RP, Kim SO. The transcription factor PU.1 mediates enhancer-promoter looping that is required for IL-1β eRNA and mRNA transcription in mouse melanoma and macrophage cell lines. *J Biol Chem*. 2019;294(46):17487-17500.

78. Gabriele M, Brandão HB, Grosse-Holz S, et al. Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science*. 2022;376(6592):496-501.

79. Mach P, Kos PI, Zhan Y, et al. Cohesin and CTCF control the dynamics of chromosome folding. *Nat Genet*. 2022;54(12):1907-1918.

80. Davidson IF, Barth R, Zaczek M, et al. CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion. *Nature*. 2023;616(7958):822-827.

81. Qi Q, Cheng L, Tang X, et al. Dynamic CTCF binding directly mediates interactions among cis-regulatory elements essential for hematopoiesis. *Blood*. 2021;137(10):1327-1339.

82. Wang H, Maurano MT, Qu H, et al. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res*. 2012;22(9):1680-1688.

83. Dubois-Chevalier J, Oger F, Dehondt H, et al. A dynamic CTCF chromatin binding landscape promotes DNA hydroxymethylation and transcriptional induction of adipocyte differentiation. *Nucleic Acids Res*. 2014;42(17):10943-10959.

84. Whalen S, Truty RM, Pollard KS. Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat Genet*. 2016;48(5):488-496.

85. Cuadrado A, Giménez-Llorente D, De Koninck M, et al. Contribution of variant subunits and associated factors to genome-wide distribution and dynamics of cohesin. *Epigenetics Chromatin*. 2022;15(1):37.

86. Kojic A, Cuadrado A, De Koninck M, et al. Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization. *Nat Struct Mol Biol*. 2018;25(6):496-504.

87. Yan J, Chen SAA, Local A, et al. Histone H3 lysine 4 monomethylation modulates long-range chromatin interactions at enhancers. *Cell Res*. 2018;28(2):204-220.

88. Sungalee S, Liu Y, Lambuta RA, et al. Histone acetylation dynamics modulates chromatin conformation and allele-specific interactions at oncogenic loci. *Nat Genet*. 2021;53(5):650-662.

89. Thiecke MJ, Wutz G, Muhar M, et al. Cohesin-Dependent and -Independent Mechanisms

Mediate Chromosomal Contacts between Promoters and Enhancers. *Cell Rep*. 2020;32(3):107929.

90. Kubo N, Ishii H, Xiong X, et al. Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. *Nat Struct Mol Biol*. 2021;28(2):152-161.

91. Hsieh THS, Cattoglio C, Slobodyanyuk E, Hansen AS, Darzacq X, Tjian R. Enhancer-promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. *Nat Genet*. 2022;54(12):1919-1932.

92. Kane L, Williamson I, Flyamer IM, et al. Cohesin is required for long-range enhancer action at the Shh locus. *Nat Struct Mol Biol*. 2022;29(9):891-897.

93. Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA. A Phase Separation Model for Transcriptional Control. *Cell*. 2017;169(1):13-23.

94. Sabari BR, Dall'Agnese A, Boija A, et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science*. 2018;361(6400). doi:10.1126/science.aar3958

95. Monfils K, Barakat TS. Models behind the mystery of establishing enhancer-promoter interactions. *Eur J Cell Biol*. 2021;100(5-6):151170.

96. Ahn JH, Davis ES, Daugird TA, et al. Phase separation drives aberrant chromatin looping and cancer development. *Nature*. 2021;595(7868):591-595.

97. Boija A, Klein IA, Sabari BR, et al. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell*. 2018;175(7):1842-1855.e16.

98. Christou-Kent M, Cuartero S, Garcia-Cabau C, et al. CEBPA phase separation links transcriptional activity and 3D chromatin hubs. *Cell Rep*. 2023;42(8). doi:10.1016/j.celrep.2023.112897

99. Ramasamy S, Aljahani A, Karpinska MA, Cao TBN, Cruz JN, Oudelaar AM. The Mediator complex regulates enhancer-promoter interactions. *bioRxiv*. Published online June 15, 2022. doi:10.1101/2022.06.15.496245

100. El Khattabi L, Zhao H, Kalchschmidt J, et al. A Pliable Mediator Acts as a Functional Rather Than an Architectural Bridge between Promoters and Enhancers. *Cell*. 2019;178(5):1145-1158.e20.

101. Shi Y, Chen J, Zeng WJ, et al. Formation of nuclear condensates by the Mediator complex subunit Med15 in mammalian cells. *BMC Biol*. 2021;19(1):245.

102. Zamudio AV, Dall'Agnese A, Henninger JE, et al. Mediator Condensates Localize Signaling

Factors to Key Cell Identity Genes. *Mol Cell*. 2019;76(5):753-766.e6.

103. van Steensel B, Furlong EEM. The role of transcription in shaping the spatial organization of the genome. *Nat Rev Mol Cell Biol*. 2019;20(6):327-337.

104. Reed KSM, Davis ES, Bond ML, et al. Temporal analysis suggests a reciprocal relationship between 3D chromatin structure and transcription. *Cell Rep*. 2022;41(5):111567.

105. Barshad G, Lewis JJ, Chivu AG, et al. RNA polymerase II dynamics shape enhancer-promoter interactions. *Nat Genet*. Published online July 10, 2023. doi:10.1038/s41588-023-01442-7

106. Zhang S, Übelmesser N, Barbieri M, Papantonis A. Enhancer-promoter contact formation requires RNAPII and antagonizes loop extrusion. *Nat Genet*. 2023;55(5):832-840.

107. Tang Z, Luo OJ, Li X, et al. CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell*. 2015;163(7):1611-1627.

108. Gaszner M, Felsenfeld G. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet*. 2006;7(9):703-713.

109. Kurukuti S, Tiwari VK, Tavoosidana G, et al. CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc Natl Acad Sci U S A*. 2006;103(28):10684-10689.

110. Arzate-Mejía RG, Josué Cerecedo-Castillo A, Guerrero G, Furlan-Magaril M, Recillas-Targa F. In situ dissection of domain boundaries affect genome topology and gene transcription in Drosophila. *Nat Commun*. 2020;11(1):894.

111. Lupiáñez DG, Kraft K, Heinrich V, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*. 2015;161(5):1012-1025.

112. Bolt CC, Lopez-Delisle L, Hintermann A, et al. Context-dependent enhancer function revealed by targeted inter-TAD relocation. *Nat Commun*. 2022;13(1):3488.

113. Richart L, Bidard FC, Margueron R. Enhancer rewiring in tumors: an opportunity for therapeutic intervention. *Oncogene*. 2021;40(20):3475-3491.

114. Chen Z, Snetkova V, Bower G, et al. Widespread increase in enhancer—promoter interactions during developmental enhancer activation in mammals. *bioRxiv*. Published online November 20, 2022. doi:10.1101/2022.11.18.516017

115. Casale FP, Rakitsch B, Lippert C, Stegle O. Efficient set tests for the genetic analysis of correlated traits. *Nat Methods*. 2015;12(8):755-758.

116. Stegle O, Parts L, Piipari M, Winn J, Durbin R. Using probabilistic estimation of expression

residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc*. 2012;7(3):500-507.

117. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575.

118. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.

119. O'Connor MF, Motivala SJ, Valladares EM, Olmstead R, Irwin MR. Sex differences in monocyte expression of IL-6: role of autonomic mechanisms. *Am J Physiol Regul Integr Comp Physiol*. 2007;293(1):R145-51.

120. Carlisle SM, Qin H, Hendrickson RC, et al. Sex-based differences in the activation of peripheral blood monocytes in early Parkinson disease. *NPJ Parkinsons Dis*. 2021;7(1):36.

121. Klein SL, Flanagan KL. Sex differences in immune responses. *Nat Rev Immunol*. 2016;16(10):626-638.

122. Lu C, Donners MMPC, Karel J, et al. Sex-specific differences in cytokine signaling pathways in circulating monocytes of cardiovascular disease patients. *Atherosclerosis*. Published online April 14, 2023. doi:10.1016/j.atherosclerosis.2023.04.005

123. Ho JSY, Mok BWY, Campisi L, et al. TOP1 inhibition therapy protects against SARS-CoV-2-induced lethal inflammation. *Cell*. 2021;184(10):2618-2632.e17.

124. Malysheva V, Ray-Jones H, Cazares TA, et al. High-resolution promoter interaction analysis in Type 3 Innate Lymphoid Cells implicates Batten Disease gene CLN3 in Crohn's Disease aetiology. *bioRxiv*. Published online November 8, 2022:2022.10.19.512842. doi:10.1101/2022.10.19.512842

125. Wingett S, Ewels P, Furlan-Magaril M, et al. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res*. 2015;4:1310.

126. Heinz S, Texari L, Hayes MGB, et al. Transcription Elongation Can Affect Genome 3D Structure. *Cell*. 2018;174(6):1522-1536.e22.

127. Yang T, Zhang F, Yardımcı GG, et al. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res*. 2017;27(11):1939-1949.

128. Kramer NE, Davis ES, Wenger CD, et al. Plotgardener: cultivating precise multi-panel figures in R. *Bioinformatics*. 2022;38(7):2042-2045.

129. Xu Z, Lee DS, Chandran S, et al. Structural variants drive context-dependent oncogene

activation in cancer. *Nature*. 2022;612(7940):564-572.

130. Corces MR, Trevino AE, Hamilton EG, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods*. 2017;14(10):959-962.

131. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol*. 2015;109:21.29.1-21.29.9.

132. Gaspar JM. NGmerge: merging paired-end reads via novel empirically-derived models of sequencing errors. *BMC Bioinformatics*. 2018;19(1):1-9.

133. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357-359.

134. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10(2):giab008.

135. Picard. Accessed July 5, 2023. http://broadinstitute.github.io/picard/

136. Ramírez F, Ryan DP, Grüning B, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016;44(W1):W160-5.

137. GitHub - ThomasCarroll/soGGi. GitHub. Accessed July 5, 2023. https://github.com/ThomasCarroll/soGGi

138. GitHub - jsh58/Genrich: Detecting sites of genomic enrichment. GitHub. Accessed July 5, 2023. https://github.com/jsh58/Genrich

139. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.

140. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48(10):1284-1287.

141. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279-1283.

142. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*. 2012;44(8):955-959.

143. Loh PR, Danecek P, Palamara PF, et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet*. 2016;48(11):1443-1448.

144. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics*. 2012;28(10):1353-1358.

145. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2013;30(7):923-930.

146. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):1-21.

147. Castel SE, Mohammadi P, Chung WK, Shen Y, Lappalainen T. Rare variant phasing and haplotypic expression from RNA sequencing with phASER. *Nat Commun*. 2016;7:12817.

148. Vigorito E, Barton A, Pitzalis C, Lewis MJ, Wallace C. BBmix: a Bayesian beta-binomial mixture model for accurate genotyping from RNA-sequencing. *Bioinformatics*. 2023;39(7):btad393.

149. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-842.

150. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079.

151. Kim-Hellmuth S, Aguet F, Oliva M, et al. Cell type-specific genetic regulation of gene expression across human tissues. *Science*. 2020;369(6509). doi:10.1126/science.aaz8528

152. Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: a tutorial (with comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors. *Stat Sci*. 1999;14(4):382-417.

153. Avsec Ž, Agarwal V, Visentin D, et al. Effective gene expression prediction from sequence by integrating long-range interactions. *Nat Methods*. 2021;18(10):1196-1203.

154. Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning–based sequence model. *Nat Methods*. 2015;12(10):931-934.

155. Bentsen M, Goymann P, Schultheis H, et al. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat Commun*. 2020;11(1):1-11.

156. Fornes O, Castro-Mondragon JA, Khan A, et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2020;48(D1):D87-D92.

157. Kulakovskiy IV, Vorontsov IE, Yevshin IS, et al. HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq

analysis. *Nucleic Acids Res*. 2018;46(D1):D252-D259.

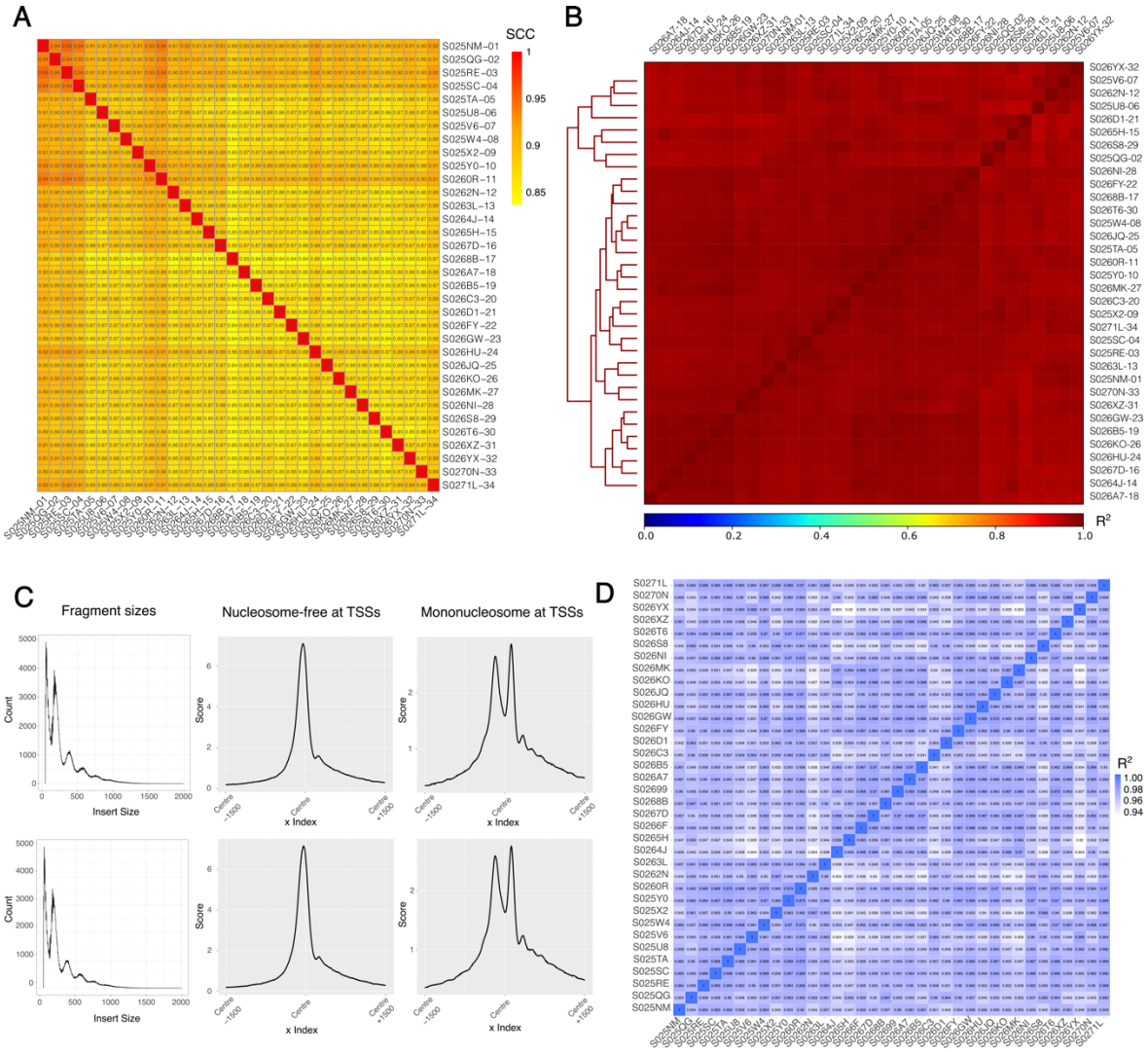
158. Coetzee SG, Coetzee GA, Hazelett DJ. motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics*. 2015;31(23):3847-3849.

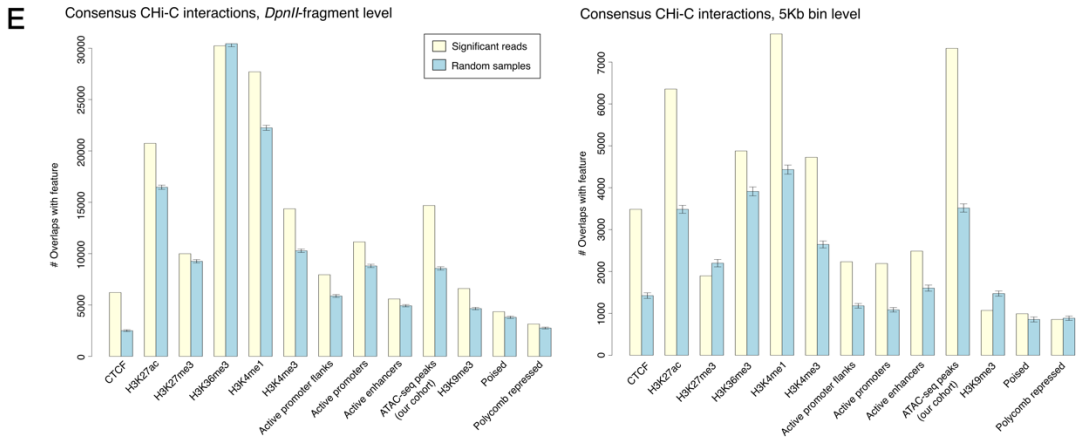
159. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1-22.

160. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.

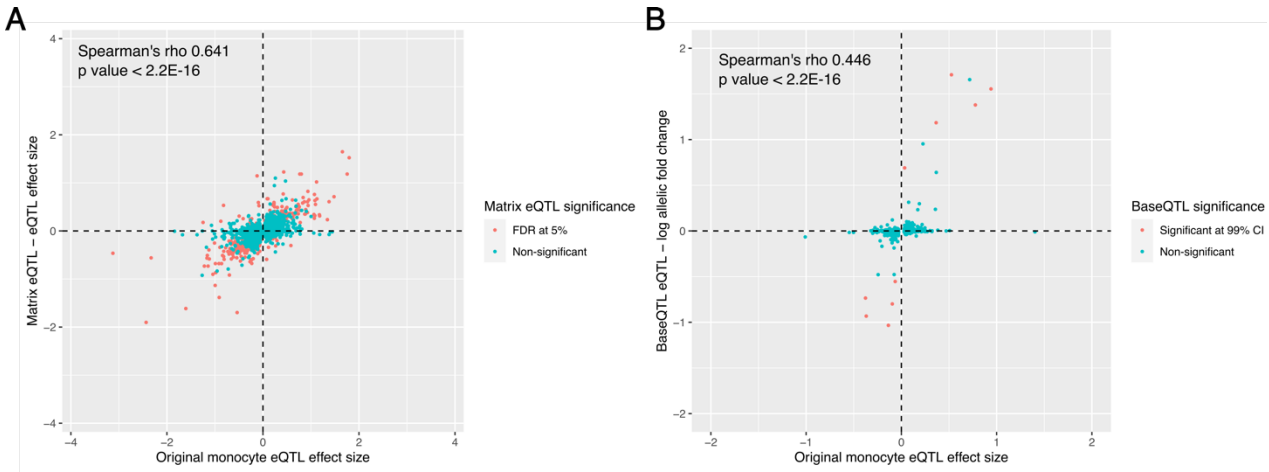
161. Sollis E, Mosaku A, Abid A, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res*. 2023;51(D1):D977-D985.

Supplementary Figures

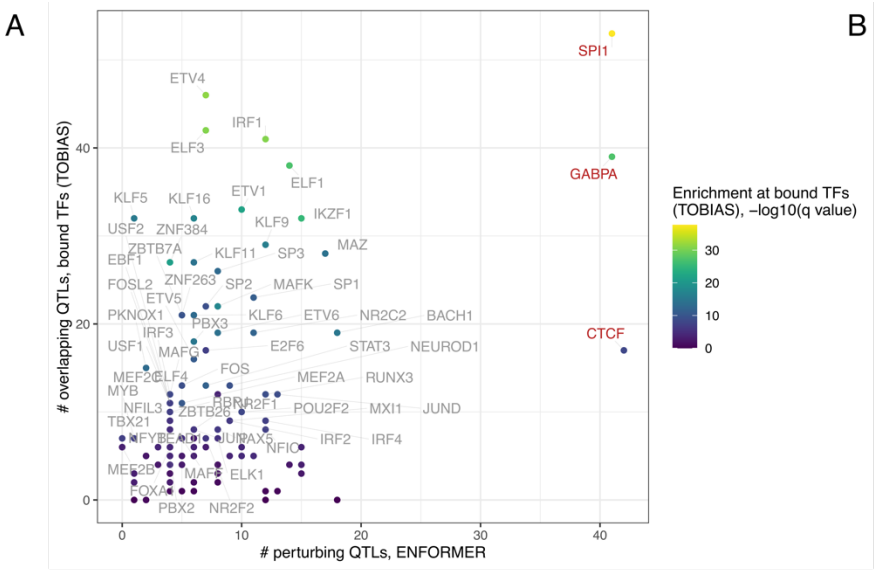




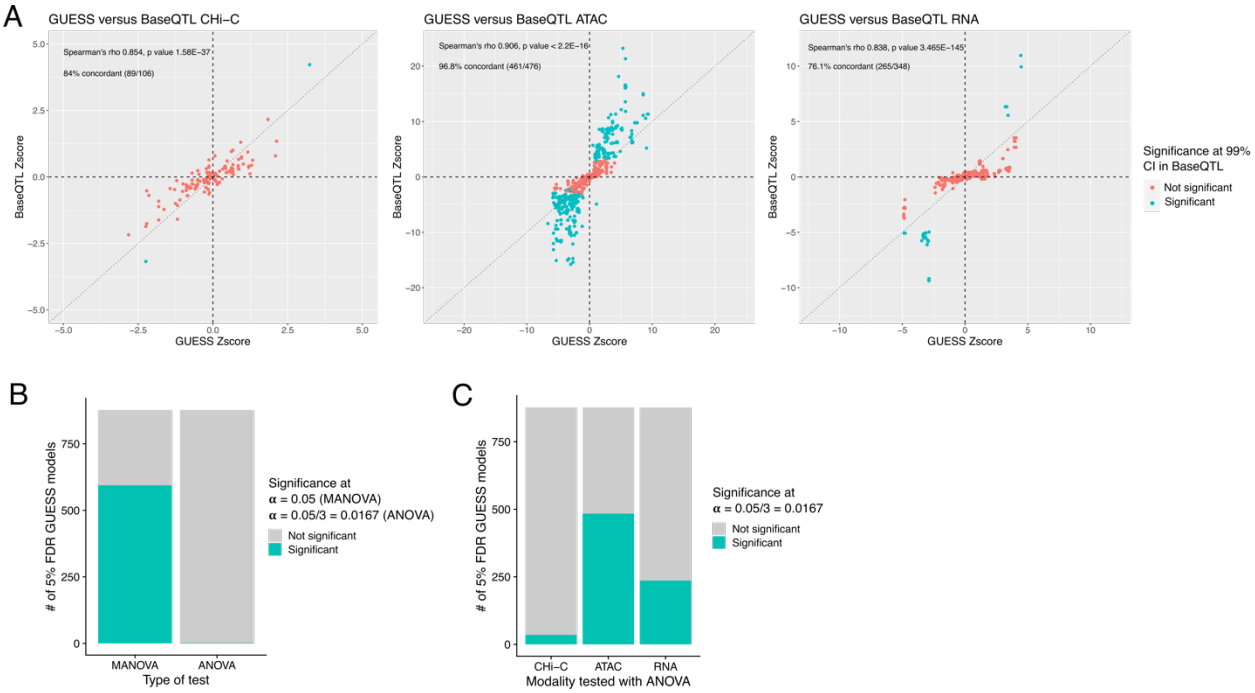
Supplementary Figure 1. QC of CHi-C, ATAC-seq and RNA-seq samples. *A) HiCRep results showing stratum-adjusted correlation coefficient (SCC) values for CHi-C data on chromosome 22. B) Pearson's correlation values for ATAC-seq samples (DeepTools). C) Representative nucleosome profiles for ATAC-seq samples. D) Pearson's correlation between RNA samples. E) CHi-C enrichments at other ends of significant interactions at DpnII-fragment level (left) and 5Kb bin level (right). Histone modification peaks were obtained from Blueprint. Annotations (promoter flanking regions, promoters, enhancers, poised and polycomb repressed regions) were taken from Blueprint projected segmentations in monocytes. ATAC-seq peaks were obtained from the present monocyte cohort (consensus HMMRATAC).*



Supplementary Figure 2. Correlation between eQTL effects in the original multi-cohort monocyte study and in the present cohort*. A) Correlation with eQTLs called by Matrix eQTL. B) Correlation with eQTLs called by BaseQTL.*


A
SPI1
GABPA
CTCF
ETV4
IRF1
ELF3
ELF1
ETV1
KLF5
KLF16
IKZF1
KLF9
USF2
ZNF384
MAZ
ZBTB7A
KLF11
SP3
EBF1
FOSL2
ZNF263
SP2
MAFK
SP1
ETV5
PKNOX1
KLF6
ETV6
NR2C2
BACH1
PBX3
IRF3
USF1
MAFG
E2F6
STAT3
NEUROD1
FOS
MEF2A
RUNX3
MYB
NFIL3
NR2F1
POU2F2
MXI1
JUND
TBX21
IRF2
IRF4
NFIC
ELK1
PBX2
NR2F2
overlapping QTLs, bound TFs (TOBIAS)
perturbing QTLs, ENFORMER
Enrichment at bound TFs
(TOBIAS), −log10(q value)
B
C
Motif match
ChIP peak
Both
Neither
E(P/NP)
CTCF
GABPA
SPI1
2.1x
5.8x
3.7x
Number of TF-perturbing cQTLs with features within a 256-bp window
D
cQTLs predicted to
perturb GABPA binding
jointly with other TFs
cQTLs
Dsrupted motifs of other perturbed TFs
FOS
FOSL2
JUND
SPI1
ATF4
CEBPB
JUN
RUNX3
ATF1
ATF3
BACH1
CREB1
CREM
CTCF
ETS1
MAFG
MAFK
TAF1
cQTLs predicted to
perturb SPI1 binding
jointly with other TFs
cQTLs
Dsrupted motifs of other perturbed TFs
GABPA
ELF1
ETS1
FOS
FOSL2
JUND
ATF4
CEBPB
CTCF
ELK1
ETV1
JUN
RUNX3
E
Chr4
SPI1
IRF8
rs881382
F
Effect size
(cQTLs vs ATAC-eQTLs)
ZNF639
MYC
KLF9
IRF1
GABPA
CEBPB
NR2F6
SPI1
BCL11A
SRF
MEF2A
EGR2
Predicted perturbed TF

Supplementary Figure 3. Predictions of cQTLs on TF binding. *A) Comparison between the number of cQTLs perturbing TFs, predicted by Enfomer, and overlapping footprints with bound TFs, detected by TOBIAS. The points are coloured by the level of q significance (Remapenrich). TFs that were significantly enriched at the cQTLs at 5% FDR (q significance > 5) are labelled. The top three perturbed TFs are labelled in red. B) Hierarchical clustering of TFs whose predicted binding was co-perturbed. C) The presence of TFBS motif matches (blue), ChIP peaks (green) or both features (pink) in the 256-bp vicinity (128 bp each way) of cQTLs predicted to perturb the binding of CTCF, GABPA or SPI1. E(P/NP): fold enrichment of the proportion of TF-perturbing cQTLs with either feature in the vicinity over the same for cQTLs not predicted to perturb the respective TF. D) Motifs perturbed by cQTLs that were associated with differential binding of GABPA (left) or SPI1 (right), but did not disrupt their cognate motifs. E) Example of a cQTL perturbing two overlapping TF motifs: SPI1 and IRF8. F) Preferential binding of TFs at cQTLs or ATAC QTLs.*



Supplementary Figure 4. Validation of GUESS results. *A). GUESS and BaseQTL concordance. The GUESS data (FDR 5%, single QTL models) and BaseQTL data (all results) were merged on the trimodal QTL and the feature: CHi-C bait fragment (left), ATAC-seq peak (middle) or gene (right). The results are coloured by their significance in BaseQTL at a CI cutoff of 0.99. Only single-QTL models in GUESS were used for the comparisons in this figure. Concordance is defined as a QTL having the same direction of effect across analyses. B) We performed multivariate analysis of variance (MANOVA) on the lead trimodal QTL identified by GUESS in each window, treating the other trimodal QTLs in the same window (if detected at 5% FDR) as covariates (MANCOVA), with the graph showing the number of models at alpha = 0.05. We also assessed how many of the lead trimodal QTLs could be validated using a univariate approach (ANOVA/ANCOVA), with the graph showing the number of models at alpha = 0.0167 for all three modalities (adjusting for multiple testing). C) Validation of each of the three individual modalities using ANOVA/ANCOVA (alpha = 0.0167).*