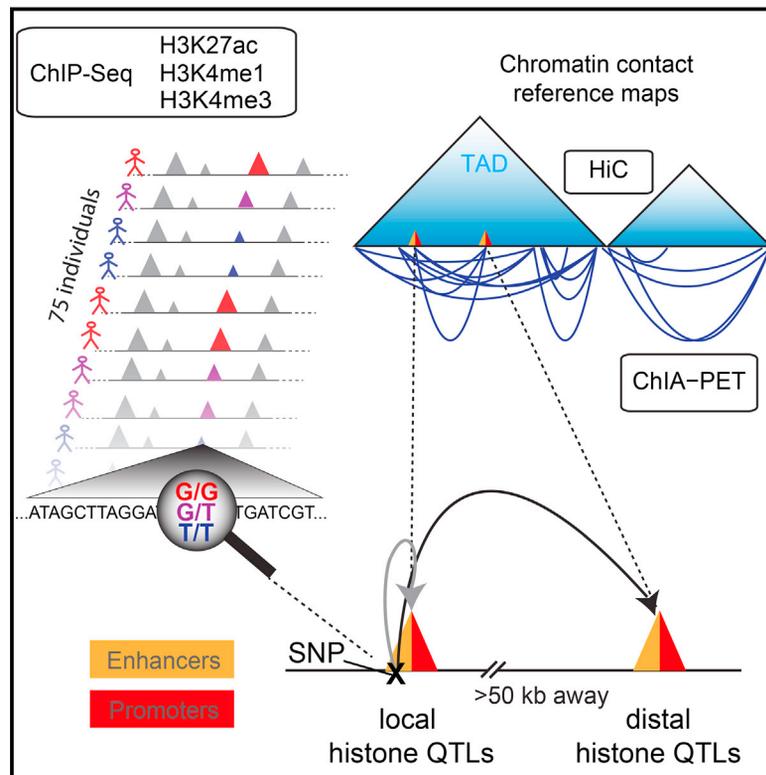


Genetic Control of Chromatin States in Humans Involves Local and Distal Chromosomal Interactions

Graphical Abstract



Authors

Fabian Grubert, Judith B. Zaugg, Maya Kasowski, ..., Lars M. Steinmetz, Anshul Kundaje, Michael Snyder

Correspondence

mpsnyder@stanford.edu

In Brief

Genetic variation in regulatory elements can effect coordinate changes in chromatin state and gene expression at both local and distal sites, reflecting associations in a three-dimensional context. Integrating information from expression, chromatin modification, and chromosome contact analyses provides a framework for assessing disease-associated mutations.

Highlights

- Analyses of variations in histone marks reveal histone QTLs in regulatory elements
- Physically interacting loci show genetically coordinated chromatin levels
- Regulatory elements sharing hQTLs are enriched in topologically associated domains
- hQTLs are enriched for GWAS SNPs and enable identification of putative target genes

Accession Numbers

GSE62742

Genetic Control of Chromatin States in Humans Involves Local and Distal Chromosomal Interactions

Fabian Grubert,^{1,7} Judith B. Zaugg,^{1,2,7} Maya Kasowski,^{1,7} Oana Ursu,^{1,7} Damek V. Spacek,¹ Alicia R. Martin,¹ Peyton Greenside,³ Rohith Srivas,¹ Doug H. Phanstiel,¹ Aleksandra Pekowska,² Nastaran Heidari,¹ Ghia Euskirchen,¹ Wolfgang Huber,² Jonathan K. Pritchard,^{1,4,5} Carlos D. Bustamante,¹ Lars M. Steinmetz,^{1,2} Anshul Kundaje,^{1,6} and Michael Snyder^{1,*}

¹Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA

²The European Molecular Biology Laboratory Heidelberg, 69117 Heidelberg, Germany

³Biomedical Informatics Graduate Training Program, Stanford University School of Medicine, Stanford, CA 94305, USA

⁴Department of Biology, Stanford University, Stanford, CA 94305, USA

⁵Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA

⁶Department of Computer Science, Stanford University, Stanford, CA 94305, USA

⁷Co-first author

*Correspondence: mepsnyder@stanford.edu

<http://dx.doi.org/10.1016/j.cell.2015.07.048>

SUMMARY

Deciphering the impact of genetic variants on gene regulation is fundamental to understanding human disease. Although gene regulation often involves long-range interactions, it is unknown to what extent non-coding genetic variants influence distal molecular phenotypes. Here, we integrate chromatin profiling for three histone marks in lymphoblastoid cell lines (LCLs) from 75 sequenced individuals with LCL-specific Hi-C and ChIA-PET-based chromatin contact maps to uncover one of the largest collections of local and distal histone quantitative trait loci (hQTLs). Distal QTLs are enriched within topologically associated domains and exhibit largely concordant variation of chromatin state coordinated by proximal and distal non-coding genetic variants. Histone QTLs are enriched for common variants associated with autoimmune diseases and enable identification of putative target genes of disease-associated variants from genome-wide association studies. These analyses provide insights into how genetic variation can affect human disease phenotypes by coordinated changes in chromatin at interacting regulatory elements.

INTRODUCTION

Deciphering the genetic and molecular basis of human traits and disease is a fundamental problem in biology and personalized medicine. Genome-wide association studies (GWAS) and deep sequencing efforts have identified common and rare single nucleotide genetic variants, as well as structural variants associated with diseases ranging from inflammatory bowel disease and Alzheimer's disease to cancer (Hindorf et al., 2009; Lambert et al., 2013; Rivas et al., 2011; Zuk et al., 2014). A majority of disease-associated common variants lie in poorly annotated non-coding genomic regions. Identifying target genes of these variants is

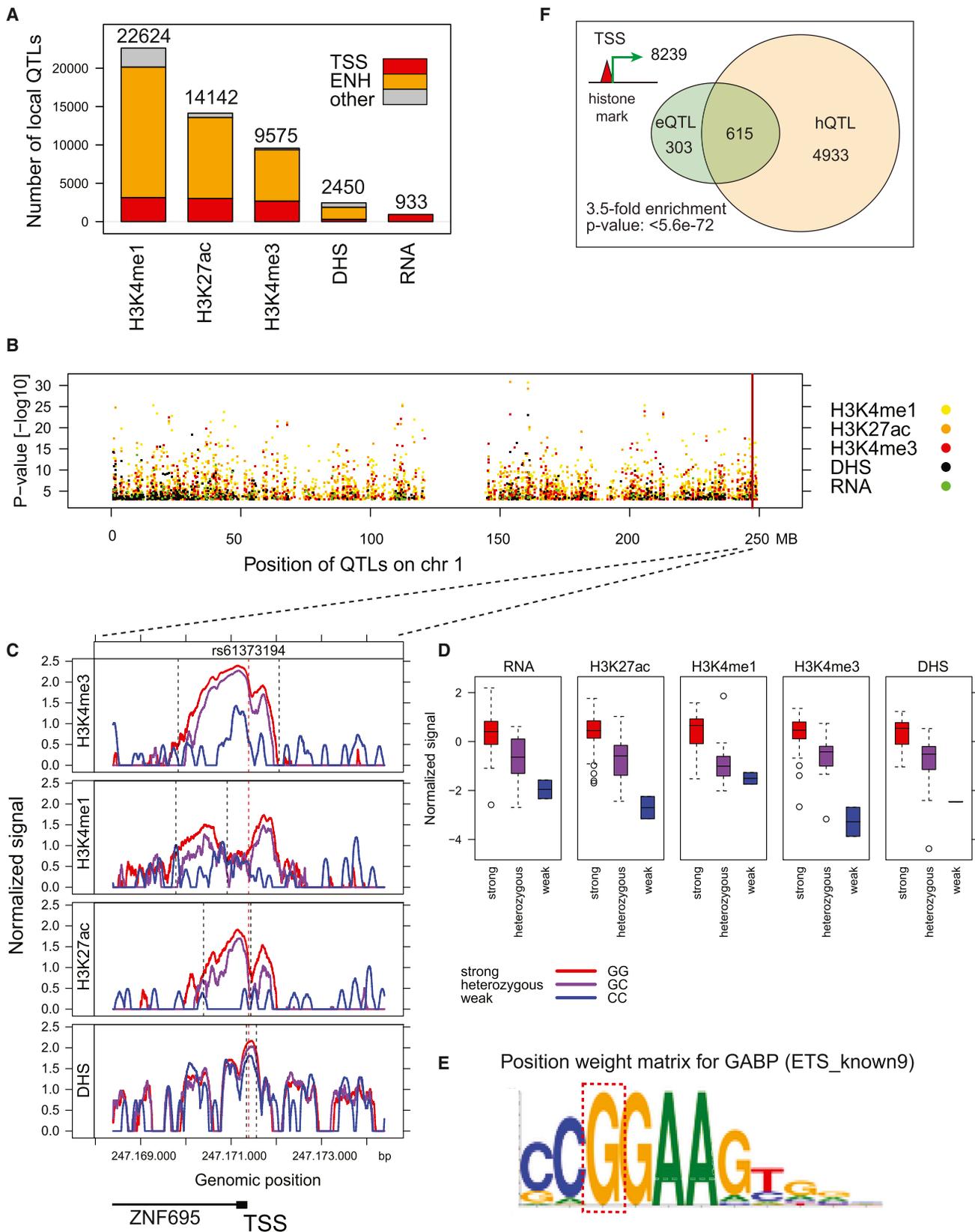
complicated by the fact that many of these variants lie in distal enhancers (Harismendy et al., 2011; Maurano et al., 2012; Pomerantz et al., 2009; Schaub et al., 2012; Smemo et al., 2014), which can regulate genes several megabases (Mbs) away through long-range chromatin contacts (Kleinjan and van Heyningen, 2005; Nobrega et al., 2003; Sanyal et al., 2012; Visel et al., 2009).

In this study, we integrated long-range chromatin contact maps with genetic variation and multiple molecular phenotypes to elucidate a comprehensive regulatory network of local and distal-acting regulatory and expression quantitative trait loci (QTLs). We mapped inter-individual variation of three histone marks associated with regulatory elements such as enhancers and promoters in lymphoblastoid cell lines (LCLs) from a cohort of 75 individuals and used these to identify histone QTLs (hQTLs). In addition, we generated Hi-C and chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) chromatin contact maps (Fullwood et al., 2009; Lieberman-Aiden et al., 2009) and used them to identify genetic variants that act over large genomic distances potentially through long-range interactions. We identified hQTLs for ~10% of predicted regulatory elements in LCLs. Fifteen percent of those were also associated with chromatin state changes at distal elements. Our findings indicate that genetic variation in regulatory elements can act over large distances to affect concordant changes in chromatin states and expression at distal sites often through transcription factor (TF) motif disruptions. hQTLs in LCLs were enriched in GWAS SNPs associated with immune-mediated diseases allowing us to predict putative target regulatory elements and genes of several disease-associated common variants. Overall, our study provides novel insights into how genetic variation can affect human disease phenotypes by disrupting regulatory networks mediated by long-range chromatin interactions.

RESULTS

Profiling Chromatin Variation in a Large Cohort of Lymphoblastoid Lines Enables Comprehensive Mapping of Local Histone QTLs

To study the relationship between genetic variation and chromatin activity, we generated chromatin immunoprecipitation



(legend on next page)

sequencing (ChIP-seq) data for three histone marks (H3K4me3, H3K4me1, and H3K27ac) in lymphoblastoid cell-lines (LCLs) from 75 unrelated individuals of the Yoruba (YRI) population, for whom high quality genotypes were available (1000 Genomes Project). H3K4me3 is primarily associated with promoters, H3K4me1 with active, bivalent, and weak enhancers, and H3K27ac with active promoters and enhancers (Ernst et al., 2011). We obtained ~25 million uniquely mapping paired-end reads (2 × 101 bp) per experiment (Figures S1A and S1B; Table S1; Supplemental Experimental Procedures). In addition, we utilized DNase-I hypersensitivity (DHS) data, which assays chromatin accessibility (for 68 YRI individuals) (Degner et al., 2012) and RNA sequencing (RNA-seq) data to quantify gene expression (for 54 YRI individuals) (Lappalainen et al., 2013).

To identify single nucleotide polymorphisms (SNPs) associated with chromatin variation, we performed quantitative trait locus (QTL) mapping. We used the ChIP-seq and DHS data to identify and quantify normalized signal at regions of enrichment (peaks). We also quantified normalized expression levels for all genes using the RNA-seq data (Figures S1C and S1D; Supplemental Experimental Procedures). All signal measurements were corrected for confounding factors (Figures S1E and S1F; Supplemental Experimental Procedures). To identify local QTLs, we tested for associations of histone/DHS peaks/gene expression and the most strongly associated SNP within ± 2 kb of peak boundaries/promoters using a linear regression framework (Degner et al., 2012), resulting in a total of 22,624, 14,142, and 9,575 local hQTLs (10% false discovery rate [FDR]) linked to 10%–15% of H3K4me1, H3K27ac, and H3K4me3 peaks respectively, as well as 2,450 DNase I sensitivity quantitative trait loci (dsQTLs) and 933 expression quantitative trait loci (eQTLs) (Figures 1A, 1B, and S1G; Table S2; Supplemental Experimental Procedures).

74% and 19% of all local QTLs were associated with enhancer and promoter chromatin states, respectively (Supplemental Experimental Procedures). QTLs often jointly influence multiple molecular phenotypes within the same region (Figure S1H). One example is shown in Figures 1C–1E where disruption of a GABP TF binding site is associated with decreased signal of three histone marks (hQTL), chromatin accessibility (dsQTL), and RNA expression (eQTL) at the *ZNF695* gene promoter (Figures 1C–1E). Globally, we find that 66% of local eQTLs are also local hQTLs (3.5-fold-enriched, p value = $5.6e-72$, Fisher's

exact test; Figure 1F; Supplemental Experimental Procedures). eQTLs that do not coincide with hQTLs may influence RNA processing rather than steady-state gene expression (Lappalainen et al., 2013), as over 30% of them reside in the 5'UTR of at least one mRNA isoform (Supplemental Experimental Procedures).

Long-Range Chromatin Contacts Are Associated with Concordant Regulatory Variation across Distal Genomic Elements

Long-range chromatin contacts are responsible for physical interactions between distal regulatory and transcribed elements. We hypothesized that genetic variants associated with local variation in chromatin state may also correlate with the activity of physically linked distal elements. To address this, we generated Hi-C data (1.4 billion reads) in a reference LCL (GM12878) to obtain a chromatin contact matrix (resolution at single restriction fragment length, median = 2,274 bp, mean = 3,697 bp). We used a covariance measure as a robust proximity estimate to identify putatively physically interacting elements (proximity score >0.4; Figures S2A–S2E; Supplemental Experimental Procedures). We tested whether a SNP is more likely to affect a coordinated gain or loss of histone modification, DHS, or RNA signal at pairs of local (<2 kb from SNP) and distal (>50 kb) elements if the pairs showed evidence of physical interaction (Figure 2A). Consistent with our hypothesis, pairs of physically interacting peaks, or genes, showed more concordant associations of SNPs with local and distal chromatin variation (Figures 2B, 2C, and S2F).

Since regions that are proximal in linear genomic distance are also more likely to physically interact, Hi-C supported SNP-peak pairs span shorter genomic distances than non-interacting pairs. To account for this distance bias, we calculated the enrichment of Hi-C interactions linking local-distal pairs that showed a genetic association (p value < 10^{-6}) for increasing distance cut-offs (Supplemental Experimental Procedures). As distance increases between a SNP and a distal element, Hi-C-linked pairs are increasingly enriched for genetic associations (Figure 2D). The effect is particularly strong for RNA expression (>10-fold enrichment of genetic associations in Hi-C connected fragments at distances >200 kb). Interestingly, physically interacting regions involving distal DHS peaks are the exception to this observation (Figures 2C and S2F). This could be explained by the largely local nature of DHS regulation in contrast to more distally

Figure 1. Local QTLs

- (A) Number of local QTLs (10% FDR) for histone marks, RNA expression (Lappalainen et al., 2013), and DHS sites (Degner et al., 2012). Peaks are classified as promoters ("TSS"), "enhancer (ENH)," and "other" based on previously defined chromatin states (Kasowski et al., 2013).
- (B) Distribution of local QTLs on chromosome 1 for histone marks, DHSs, and RNA. The vertical red line marks the location of the local joint QTL described in (C)–(E).
- (C) Signal tracks for three histone marks and DHS in a 5 kb region around a joint local hQTL/dsQTL/eQTL coinciding with the *ZNF695* promoter. The signal is aggregated across individuals by their genotype at rs61373194. The position of the QTL SNP is indicated with a vertical dashed red line.
- (D) Boxplots of aggregated signal for gene expression, histone marks, and DHS grouped by the genotype of the QTL SNP. The normalized signal corresponds to the signal averaged across the entire peak region as indicated by black dashed lines in (C).
- (E) Position weight matrix (PWM) for GABP (ETS_known9). The third position corresponds to the location of the motif altering SNP (rs61373194). The genotype with the strongest signal corresponds to a better match to the consensus sequence. Note: the motif is on the (–) strand. Therefore the genotypes are shown for the (–) strand to correspond to the PWM.
- (F) Overlap between local eQTLs and hQTLs (calculated for promoters within 5 kb of a histone peak). The overlap is highly significant (Fisher's exact test); two-thirds of all eQTLs are also hQTLs. There are 8,239 promoters that coincide with non-QTL histone peaks.
- See also Figure S1 and Tables S1 and S2.

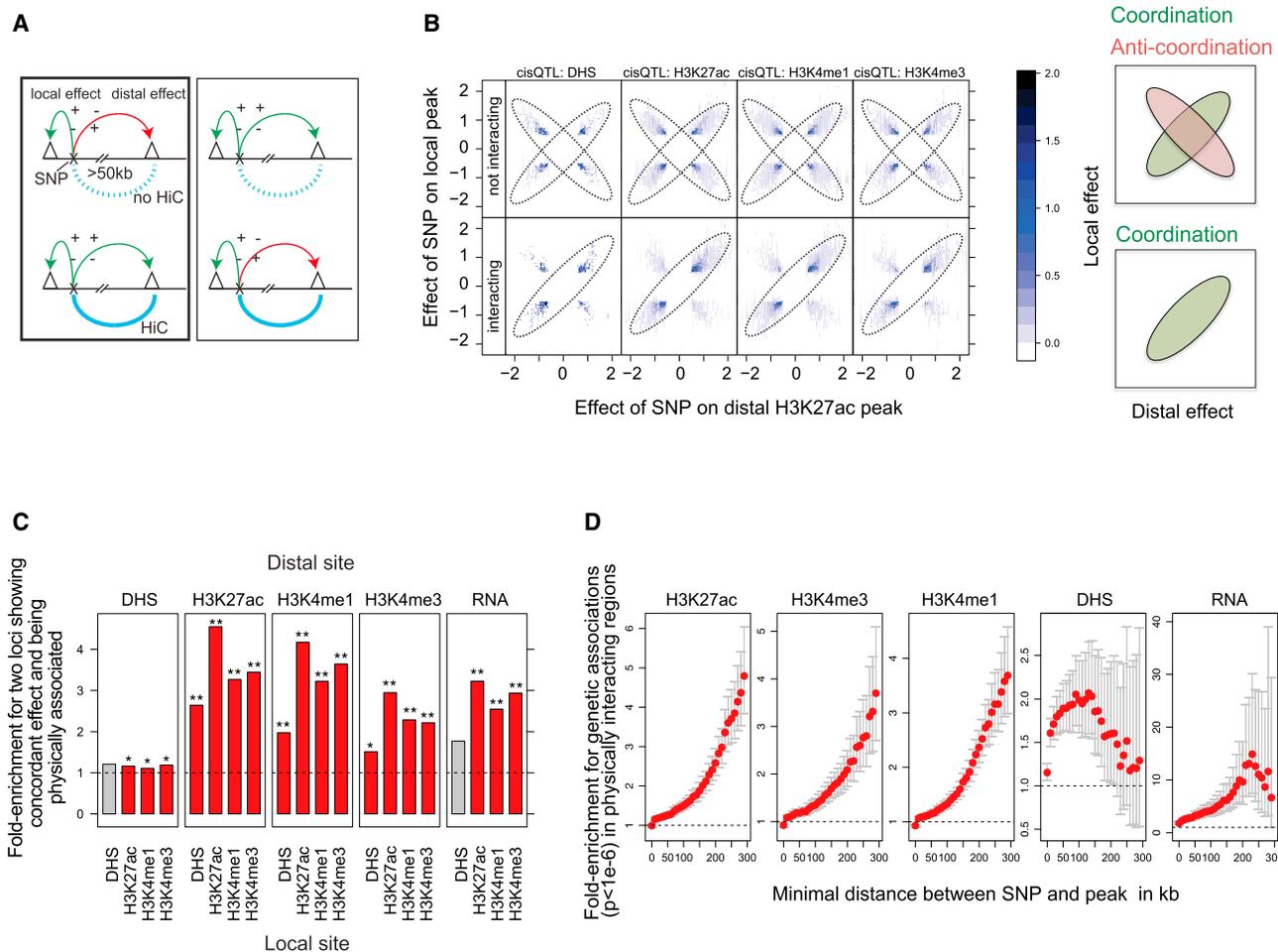


Figure 2. Evidence of Genetic Coordination among Distal Functional Elements

(A) Schematic of the four possibilities for genetic and physical coordination: the effect of a local QTL on a distal peak can be in the same (+/+ and -/-) or opposite direction (+/- and -/+) and the local QTL can be physically interacting with distal peak or not (bottom versus top).

(B) The effect (β of the regression) of a local QTL on its local peak (y axis) and distal H3K27ac peak (x axis) grouped by Hi-C correlation score between both loci into "interacting" (>0.4) and "non interacting" (≤ 0.4) (bottom versus top). The direction of the effect tends to be in the same direction for physically interacting pairs but not for non-interacting pairs (see also [Figure S2F](#) for QTLs distal to the other marks, DHSs, and RNA).

(C) Enrichment of physical interaction among pairs of local QTLs and all distal genomic features (>50 kb) that covary in the same direction with respect to the local QTL SNP. The enrichment is shown for all combinations of histone marks, RNA, and DHSs (Fisher's exact test, red bars indicate significance with * $p < 0.01$ and ** $p < 10^{-10}$).

(D) Enrichment for genetic associations (p value $< 10^{-6}$) in physically interacting regions as a function of the minimal distance between the SNP-distal peak pair. Enrichments and 95% confidence intervals are calculated using Fisher's exact test.

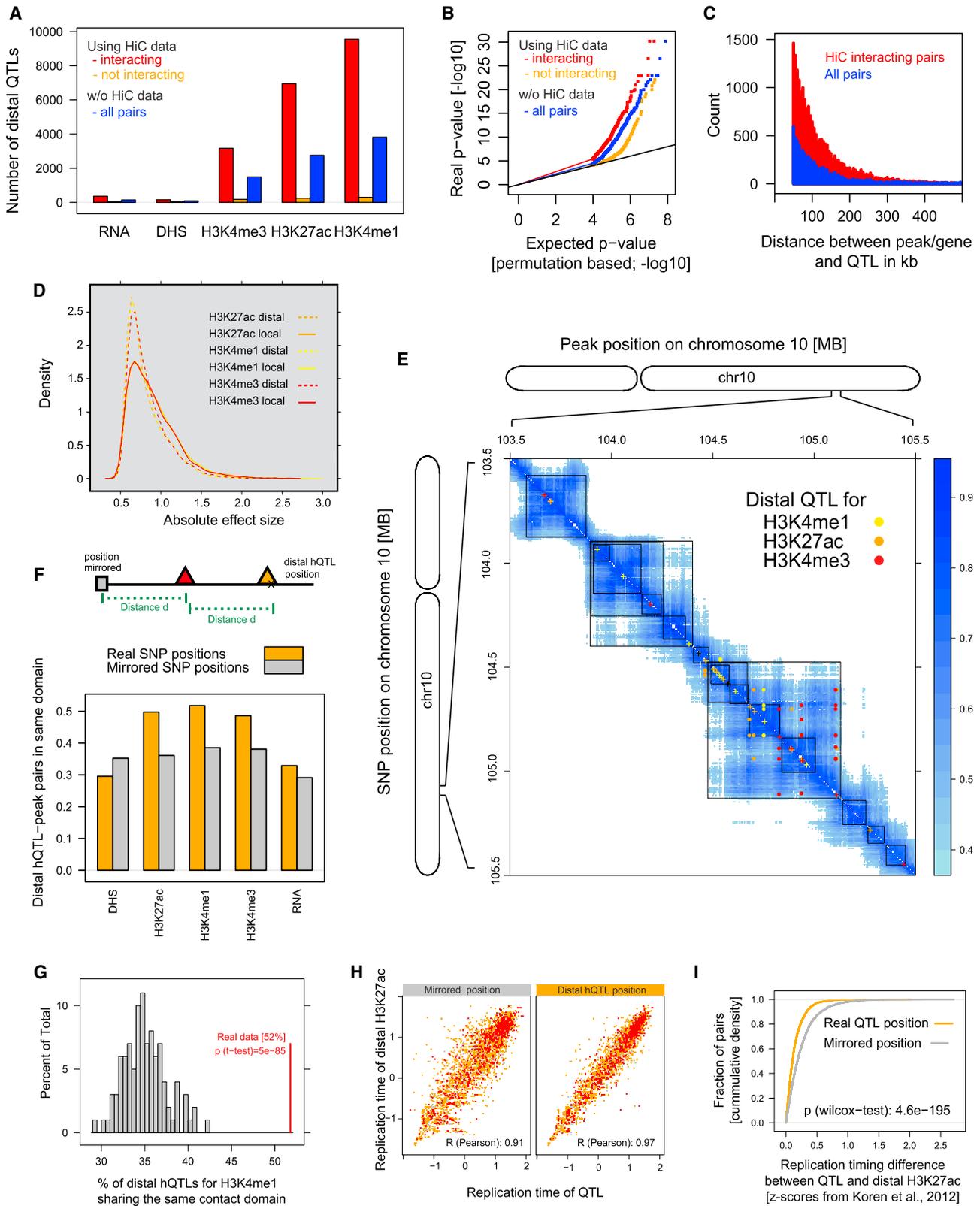
coordinated regulatory changes in histone marks, or alternatively it could simply be due to the greater power of our deeply sequenced histone data.

Long-Range Chromatin Contact Maps Increase Power for Distal QTL Discovery

Given the support for genetically coordinated regulatory variation through long-range chromatin contacts, we next sought to identify local QTLs associated with physically interacting distal (>50 kb) molecular phenotypes ("distal QTLs"). We paired SNPs that were local h/ds/eQTLs with distal (within 2 Mb) histone/DHS peaks and genes and grouped the pairs into whether or not they are physically interacting based on Hi-C proximity scores. We then as-

sessed the significance of the associations independently for both groups, leveraging the Hi-C data as an orthogonal filter ([Bourgon et al., 2010](#)). To ensure that our distal QTLs were independent (i.e., not in linkage-disequilibrium (LD) with a local QTL for the same peak) we (1) identified and kept the most significant SNP (regardless of distance) for each peak; (2) removed all SNPs in LD ($r^2 > 0.2$) with the SNP identified in the previous step; repeated steps (1) and (2) until no SNPs were remaining ([Figure S3D](#); [Supplemental Experimental Procedures](#)). From this set of independent QTLs for each peak/gene, we defined distal QTLs as being located >50 kb from the peak/gene ([Table S2](#)).

The resulting QTLs were compared to those obtained from a standard QTL analysis. Using Hi-C as a filter, we identified



(legend on next page)

20,950 distal QTLs across all tested molecular phenotypes as compared to 8,305 using the standard approach (10% FDR) (Figures 3A and S3A; Supplemental Experimental Procedures). Hi-C data increases power to detect distal QTLs irrespective of genomic distance constraints with increased gains for larger window sizes by focusing the association tests on SNP-peak pairs that are more likely to be associated (Figure 3B; control of permuted Hi-C links, Figures S3B and S3C; Supplemental Experimental Procedures). Most distal QTLs were located within 200 kb from the associated functional element (Figure 3C). We hypothesized that if QTLs can affect regulatory variation at distal elements through transient chromatin contacts, it is likely that a variant will have a weaker effect on a distal element compared to a local element. Consistent with this hypothesis, we find smaller effect sizes for distal QTLs compared to local QTLs (p values $< 10^{-16}$; Wilcoxon rank sum test) (Figure 3D).

Distal Chromatin QTLs Are Enriched within Topologically Associated Domains

Distal QTLs for multiple molecular phenotypes were found to often fall in regions with high density of physical interactions (Figure 3E) that highly overlap with topologically-associated domains (TADs) (Dixon et al., 2012; Rao et al., 2014) (Figure S2E). Mirroring the real position of the SNP relative to the distal molecular trait produced fewer pairs residing within TADs (Figure 3F), suggesting that distal chromatin QTLs are likely to be enriched within TADs. To explicitly test this hypothesis, we shuffled the locations of TADs and compared the number of distal QTLs falling within the actual TADs versus numerous shuffled sets (Supplemental Experimental Procedures). We observed a significant enrichment (p value $< 10^{-70}$; one-sided t test) of distal QTLs involving all three histone marks (Figures 3G and S3E) and a corresponding lack of enrichment for distal associations across TAD boundaries (Figure S3F). This result was confirmed in an

alternate analysis in which we examined the fraction of distal QTLs falling within TADs compared to random sets of distance-matched SNP-peak pairs. Again, distal QTLs for all three histone marks, but not RNA ($p = 0.19$) and DHS ($p = 0.25$), were found to be enriched within TADs (p values ≤ 0.01 ; Fisher's exact test). Similarly, the replication timing (Koren et al., 2012) of pairs of regions connected by a distal QTL were more similar compared to equidistant control regions showing no significant genetic associations (Figures 3H and 3I).

hQTLs Form a Highly Connected Network of Diverse Physically Interacting Regulatory and Transcribed Elements

To further dissect the functional elements associated with regulatory variants, we utilized combinatorial chromatin states (Kasowski et al., 2013) and GENCODE v19 transcriptome annotations (Harrow et al., 2012) to label peaks as putative enhancers or promoters (Supplemental Experimental Procedures). The majority of local-distal QTLs (88%) capture associations between and across enhancers and promoters (Figure 4A). QTLs involving enhancer-enhancer pairs represent the largest and, for the enhancer marks (H3K4me1, H3K27ac) a significantly higher than expected, proportion (54%) of distal hQTLs (Figures 4A and 4B). A total of 5.7% of QTLs involve pairs of local and distal promoters (>2 -fold more than expected for all marks and RNA; Figure 4B) consistent with Pol II ChIA-PET studies reporting extensive promoter-promoter links (Li et al., 2012).

We systematically studied the properties of genetically coordinated regulatory elements. For enhancer and promoter local hQTLs, we computed the "out-degree," i.e. the number of outgoing associations to distal enhancers and/or promoters (Figures 4C and 4D). For regulatory elements that are linked by at least one distal QTL, we computed their number of associated incoming distal hQTLs ("in-degree"; Figures S4A and S4B).

Figure 3. Distal QTLs

- (A) Number of distal QTLs identified with and without Hi-C. To identify distal QTLs we used SNPs that are a local QTL for any of the histone marks or RNA, split the set of SNP-peak pairs based on whether or not they are physically interacting (Hi-C correlation > 0.4), and calculated the FDR for each set individually (Figure S3A; Supplemental Experimental Procedures).
- (B) Q-Q plot for the different sets of SNP-peak pairs for H3K27ac. The expected p values were calculated by permuting sample labels (Supplemental Experimental Procedures). Q-Q plots for the other marks as well as a control set using permuted Hi-C links is shown in Figures S3A and S3B.
- (C) Distribution of distances between local-distal hQTL peak pairs identified using Hi-C interaction data or a standard hQTL approach. Most distal QTLs are <200 kb apart from their targets.
- (D) Distribution of absolute effect sizes of local and distal QTLs. Globally, local effects are stronger than distal effects.
- (E) Heatmap of estimated Hi-C interaction counts for a region harboring multiple significant distal and local QTLs. The Hi-C interactions are based on a covariance method (Supplemental Experimental Procedures). Hi-C fragments with a correlation score >0.4 were considered interacting. Local QTLs are indicated on the diagonal; off-diagonal dots indicate distal QTLs (>50 kb). Black squares correspond to contact domain calls from Rao et al. (2014).
- (F) The position of distal QTLs was mirrored (see schematic) to test whether local-distal QTL pairs are more likely to reside within the same TAD. The schematic indicates how QTL SNPs were mirrored relative to the associated feature (histone peak, DHS, or RNA). Shown is the fraction of feature-QTL pairs that share the same TAD (orange) compared to feature-mirrored QTL pairs (gray). Features more frequently share the same TAD with the true QTL than the mirrored QTL. To avoid any bias we used the non-Hi-C aware set of QTLs for this analysis.
- (G) Distribution of fractions of local-distal H3K4me1 QTLs sharing the same TAD for 100 sets of shuffled TADs. The fraction of QTL pairs sharing true TADs is indicated by the red vertical line. For chromatin marks, the fraction of local-distal hQTLs that share the same TAD is significantly higher than for the shuffled TADs (one-sided t test). To avoid any bias we used the non-Hi-C aware set of QTLs for this analysis. The same analysis for the other marks is shown in Figure S3E.
- (H) Comparison of replication timing for local-distal QTL pairs. Shown are scatterplots of the replication timing for an H3K27ac peak region (y axis) and the region of its distal QTL (x axis; right panel) or mirrored QTL (see schematic in Figure 3F) (x axis; left panel). The peak-real QTL pairs show higher correlation than the peak-mirrored QTL pairs (Pearson correlation).
- (I) The cumulative distribution of replication timing difference between peak and QTL regions (orange) and peak and mirrored QTL regions (gray) is shown. The differences for the hQTL-real peak region pairs are significantly smaller than for the corresponding mirrored positions (Wilcoxon rank sum test). See also Figure S3 and Table S2.

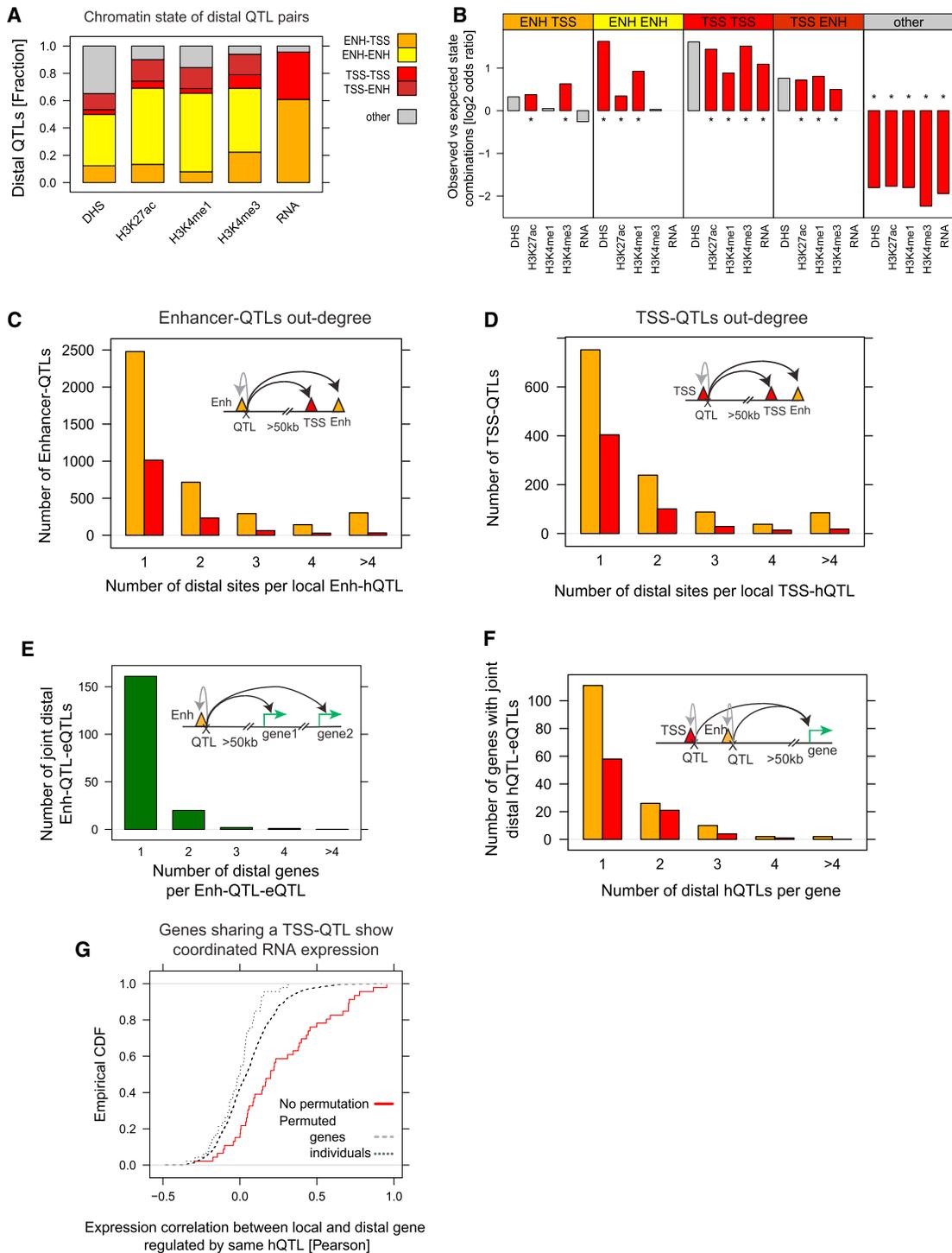


Figure 4. Characterization of Distal QTLs

(A) Fraction of chromatin states for local-distal QTL pairs grouped by mark at distal site. Most pairs involve associations between pairs of enhancers (yellow) or enhancers and promoters (TSS, orange).

(B) The log odds ratio of observed versus expected combinations of chromatin states for local and distal QTL pairs is shown for each chromatin mark, RNA, and DHS. The number of expected combinations for each mark was obtained by permuting the state labels for the local peaks. (Fishers exact test * = p value < 0.01)

(C and D) Number of local enhancer QTLs (C) and local promoter QTLs (D) influencing one or more distal promoters or enhancers (“out-degree”). In-degree plots are shown in Figure S4.

(E) Out-degree (as in C and D) of local enhancer QTLs that are also eQTLs for one or more distal genes.

(F) In-degree of genes with one or more distal enhancer/promoter.

(legend continued on next page)

QTLs local to enhancers or promoters are frequently associated with one or more distal enhancers (3,934) and promoters (1,384) (Figures 4C and 4D). We identified 184 local enhancer hQTLs with distal effects on gene expression, 12.5% of them associated with more than 1 gene (Figure 4E). Reciprocally, 26% of the 151 genes with an eQTL are associated with SNPs at multiple enhancers (Figure 4F). Interestingly, we identified 84 promoter hQTLs that are eQTLs for a distal gene (Figure 4F). As expected, these pairs of genes with a shared regulatory variant showed significantly more correlated expression than distance-matched random pairs of genes (Figure 4G; p value = $1.5e-6$, Wilcoxon rank sum test). Overall, these results suggest extensive overlap of eQTLs and hQTLs, and a subset of joint hQTL-eQTLs involves multiple regulatory elements or genes.

Looping Chromatin Contacts Connect Covarying Loci that Are Enriched for Genetic Associations

To independently assess the extent of genetic associations between physically interacting loci, we generated ChIA-PET data for the H3K4me3 promoter mark and RAD21 (a subunit of the cohesin complex involved in mediating distal chromatin contacts) in GM12878 cells (Figure 5A; Supplemental Experimental Procedures). We identified 10,220 chromatin loops for H3K4me3 and 41,532 for RAD21 (5% FDR), a large portion (~80%) were found to overlap with loops uncovered in a previous study (Rao et al., 2014) indicating the good quality of our data (Figure S5A). We characterized the loops by chromatin state at interacting sites. As expected (Jin et al., 2013; Rao et al., 2014), interactions between enhancers and promoters were highly prevalent (9,147; 24% of all ChIA-PET links), often involving multiple enhancers (2,085) linked to the same promoter (Figure 5B; Supplemental Experimental Procedures). We also find ~2,028 promoters (5% of ChIA-PET links) linked to a different distal promoter and 470 promoters linked to multiple distinct promoters (Figure 5B). H3K4me3 ChIA-PET was enriched for enhancer-promoter-interactions whereas the RAD21 ChIA-PET was enriched for promoter-promoter interactions and “other” chromatin state classes including repressed and CTCF states (Figure 5C; Fisher’s exact test).

We further used our ChIA-PET data along with Hi-C data from Rao et al. (2014) to test if physically interacting loci are enriched for genetic associations (p value $< 10^{-6}$) between SNPs and distal molecular phenotypes, as compared to random sets of distance-matched SNP-peak pairs (EEP). We observed a significant enrichment of both ChIA-PET and Hi-C loops in distal associations between SNPs and all three histone marks (not significant for H3K4me3 in Hi-C loops), as well as RNA expression levels (Figure 5D). This result was confirmed in an alternate analysis in which we shuffled the location of the loops and counted the number of associations involving SNP-peak pairs located on opposite ends of chromatin loops relative to multiple sets of shuffled loops (Supplemental Experimental Procedures). The ChIA-PET and Hi-C loops were both enriched in associa-

tions between a SNP and the H3K4me3 mark (for ChIA-PET also H3K27ac), DHS and RNA expression levels (Figure S5B), with expression showing the strongest enrichment (~2.5-fold and ~2.2-fold, in ChIA-PET and Hi-C loops, respectively).

To corroborate our observation that local-distal hQTLs tend to physically interact, we tested whether pairs of regulatory elements linked by ChIA-PET interactions also exhibit concordant variation of molecular phenotypes. Indeed, we observed significant covariation of histone/DHS signal and expression levels (Figures 5E and 5F; Wilcoxon rank sum test $p < 1e-16$) among pairs of enhancers/promoters and genes linked by ChIA-PET interactions.

hQTLs Potentially Affect Chromatin State at Regulatory Elements by Disrupting Binding Sites of Transcription Factors

To investigate the potential molecular basis of hQTLs, we used ENCODE TF ChIP-seq data from GM12878 (Dunham et al., 2012) to assess the enrichment of in vivo TF binding sites at regulatory elements associated with local hQTLs. For each TF, we computed the enrichment of overlap between histone peaks containing at least one TF peak and those associated with a local hQTL (Fisher’s exact test). While hQTL-associated H3K27ac peaks were enriched for key lymphoid TFs such as SPI1 (PU.1), BCL11A, and PAX5, hQTL-associated H3K4me3 peaks were enriched for general regulatory factors such as POLR2A, USF1, ELF1, and TAF1 (Figures 6A, 6B, S6A, and S6B). A complementary, rank-enrichment approach (Figure 6C) (Supplemental Experimental Procedures) revealed similar results: the highest scoring TFs by the Fisher’s exact test (e.g., SPI1) show significant overlap enrichments across the entire spectrum of histone ChIP-seq peaks; low scoring TFs (e.g., NRF1 and FOS) exhibit a distinct lack of enrichment at peaks showing significant QTL associations (Figure 6C).

We next tested whether disruption of TF binding motifs could explain the observed variation of histone marks. For each TF, we used one representative motif position weight matrix (PWM) (Kheradpour and Kellis, 2014). After extensive filtering to avoid false positive calls of motif matches across the genome (Supplemental Experimental Procedures), we computed the Spearman correlation between the motif PWM score at the SNP and the histone mark signal at the peak, across all individuals. For the subset of peaks/genes with a QTL that passed filtering (13% H3K4me1, 25.9% H3K4me3, 18.7% H3K27ac, 33.3% DHS, and 32.4% RNA), we found significantly correlated motifs for 89%–95% of peaks (5% FDR; Figure 6D; see Table S3 for full analysis results). For H3K27ac, TFs with the highest proportion of positive correlations include activating enhancer-associated TFs (e.g., NFkB, BCL, SPI1), whereas TFs with known repressive roles (e.g., REST, ZBTB33, SRF) show a larger proportion of negative correlations (Figure 6E; see Figures S6C–S6F for all histone marks, DHS, and RNA). Similarly, we found a correlated motif for the majority of tested distal QTL peaks (60%–73% for

(G) Pearson correlation of expression signal among pairs of distal promoters that are associated with the same hQTL (one local, one distal). The correlations are compared to a set of distance matched permuted pairs (permuted genes) and a set of real links with permuted sample names (permuted individuals). Promoters sharing an hQTL are more correlated than permuted data (Wilcoxon rank sum test, $p = 2.8e-12$). See also Figure S4.

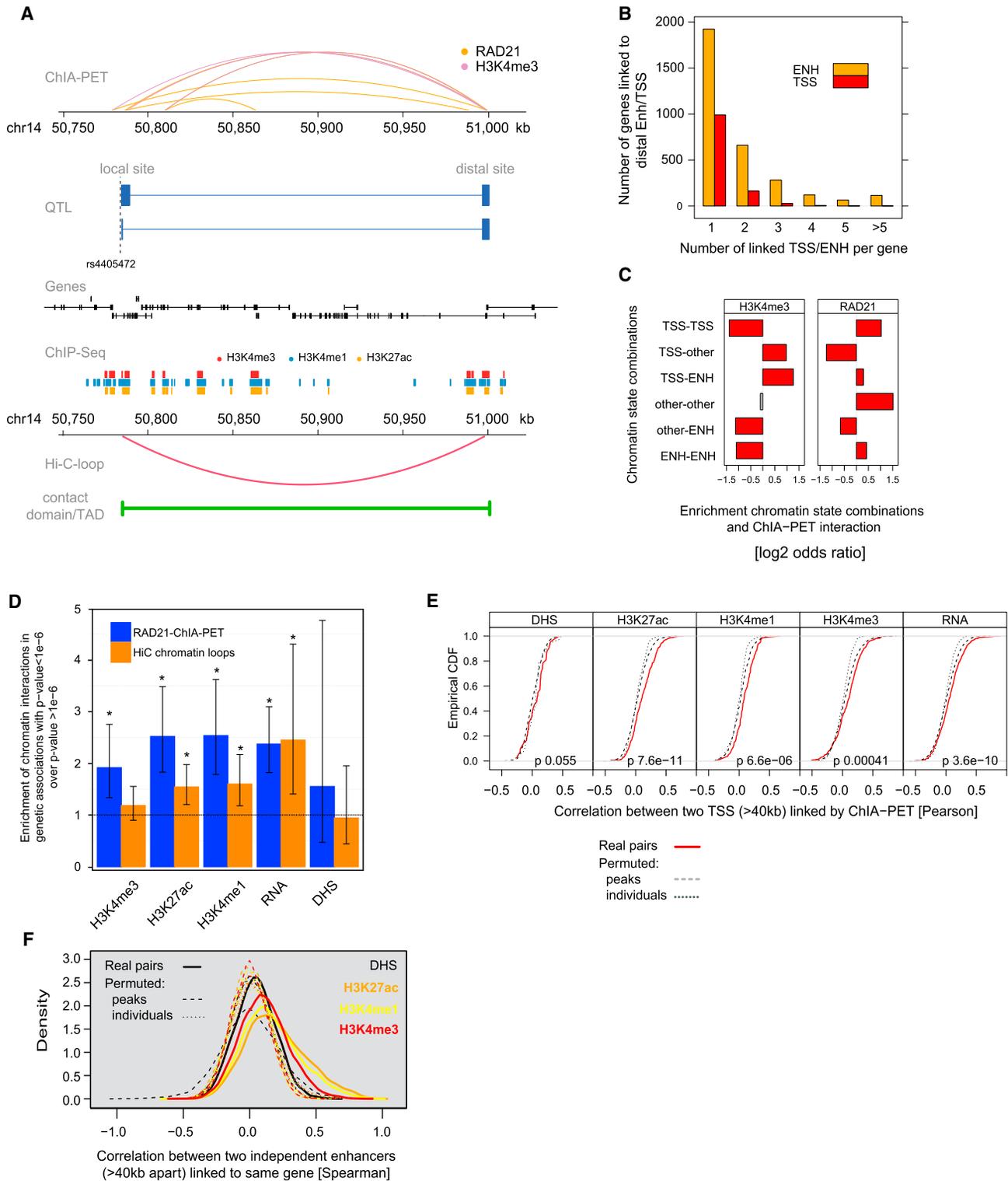


Figure 5. Chromatin Loops and ChIA-PET

(A) Example locus showing a genetic variant (rs4405472; dashed vertical line) that is a local hQTL for H3K4me1 and H3K4me3 as well as a distal hQTL for H3K4me3 at a promoter ≥ 200 kb away, skipping several genes. The two loci are physically linked as indicated by significant ChIA-PET interaction calls (top) and a Hi-C-based chromatin loop call (bottom) (Rao et al., 2014). The green line indicates a chromatin contact domain.

(B) ChIA-PET-links between one or more enhancers/promoters and distal genes based on combined interaction calls for H3K4me3 and RAD21.

(legend continued on next page)

histone marks and DHS, 35% for RNA). Joint analysis of motif correlations with local and distal peaks revealed largely concordant effects (Figure S6G), consistent with co-variation of the distally associated peaks (Figure 2B).

To further characterize the local chromatin environment of motif-disrupting quantitative trait nucleotides (QTNs), we extracted ChIP-seq signal for H3K27ac around each H3K27ac QTN (± 500 bp), averaged the signal across individuals at each location, sorted the profiles into six groups using unsupervised clustering, and aggregated the signal in each cluster by genotype (Figure 6F; Supplemental Experimental Procedures). Consistent with TF binding being a major driver of chromatin variation, we observe that the vast majority of QTNs lie in nucleosome-free regions, either between two peaks (clusters 1,2,3) or at peak shoulders (cluster 5,6) with only a small subset potentially coinciding with nucleosomes (cluster 4) (Figure 6F). Grouping and aggregating the signal for other histone marks and DHS of the same regions according to the H3K27ac clusters, we found high concordance between histone marks and DHSs (Figures 6F and S6H). Sequencing of micrococcal nuclease-digested chromatin (MNase-seq) data from GM12878 (Kundaje et al., 2012) confirmed that the chromatin mark patterns are consistent with overall nucleosome positioning (Figures 6F and S6H).

hQTLs Are Enriched for Autoimmune Disease-Associated Variants Enabling the Identification of Putative Disease-Associated Target Genes and Regulatory Networks

Enhancer elements active in disease-relevant cell types have been previously shown to be enriched for disease-associated variants from genome-wide association studies (GWAS) (Karczewski et al., 2013; Maurano et al., 2012; Nicolae et al., 2010; Schaub et al., 2012). Consistently, we found significantly more GWAS SNPs in strong LD ($r^2 \geq 0.8$) with hQTLs than SNPs matched for MAF, distance to promoter/TSS, and LD (Figure 7A; data from Table S4). When separating hQTLs into enhancers and promoters, we found a significant enrichment of GWAS SNPs for both categories in all three marks (Figures 7A and S7A). Histone peaks with hQTLs show stronger enrichment in GWAS SNPs than peaks without a hQTL (odds ratio [OR] = 1.22, $p = 6.9 \times 10^{-8}$, 95% confidence interval [CI] = [1.14, 1.32] across all SNPs in the NHGRI GWAS catalog).

Further, we found that hQTL SNPs for all histone marks are significantly enriched (Fisher's exact test) for SNPs associated (association p values $< 1 \times 10^{-5}$) with autoimmune diseases including multiple sclerosis, rheumatoid arthritis, Crohn's disease and ulcerative colitis (Figure 7B). These findings were

corroborated using rank enrichment analysis for the subset of diseases/traits for which complete summary statistics were available (Table S4; Supplemental Experimental Procedures): hQTLs were enriched in GWAS SNPs with moderate to strong associations with autoimmune diseases (Figures 7C, 7D, and S7B), but were not enriched for coronary artery disease (Figure S7B). Both methods also exhibit strong enrichment of LCL hQTLs in Alzheimer's disease (AD)-associated GWAS variants (Figures 7B and 7D), consistent with recent findings indicating that AD GWAS variants are strongly enriched in regulatory elements active in CD19⁺ primary B cells and CD14⁺ primary monocytes (Gjoneska et al., 2015).

To identify putative target genes of disease-associated hQTLs, we intersected our local and distal (GWAS were expanded using European LD, $r^2 > = 0.8$, QTLs were expanded using Yoruban LD, $r^2 > = 0.8$) hQTLs and eQTLs with genome-wide significant GWAS tag SNPs (Supplemental Experimental Procedures) for each enriched disease (Figure S7C; Table S4). Our analysis recapitulated several known target genes and provided other new candidates. For example, using Crohn's disease (CD) as a case study, we linked 115 of 425 genome-wide significant ($p < 10^{-5}$) tag SNPs to 184 regulatory elements and 33 genes (Figure 7E). We find distal hQTLs/dsQTLs for *SLC25A20*, *WDR6*, *APEH*, *RHOA*, and *TCTA* as well as distal eQTLs for *RP11-387H17.4*, *ORMDL3*, and *ERAP2*. eQTLs in LD with CD GWAS SNPs have been previously reported for the *ERAP2* and *ORMDL3* genes (Franke et al., 2010; Van Limbergen et al., 2009).

DISCUSSION

We have integrated genetic, chromatin, and expression variation data with chromatin contact maps in LCLs to characterize the effects of genetic variants on distal regulatory elements and gene expression. We developed a novel approach using three-dimensional chromatin contact maps from Hi-C data as a scaffold for long-range association testing that significantly increases power to detect distal hQTLs by restricting hypothesis testing to sites of physical proximity. Most importantly, our study provides mechanistic insights into genetically coordinated variation occurring through three-dimensional physical contacts of regulatory elements.

In utilizing three-dimensional proximity to discover QTLs, we make a number of observations pertaining to the role of chromatin structure in regulatory variation. We found that local and distal hQTLs are abundant, suggesting a strong genetic basis for the extensive variation of chromatin state (Kasowski et al., 2013; Kilpinen et al., 2013; McVicker et al., 2013). Consistent

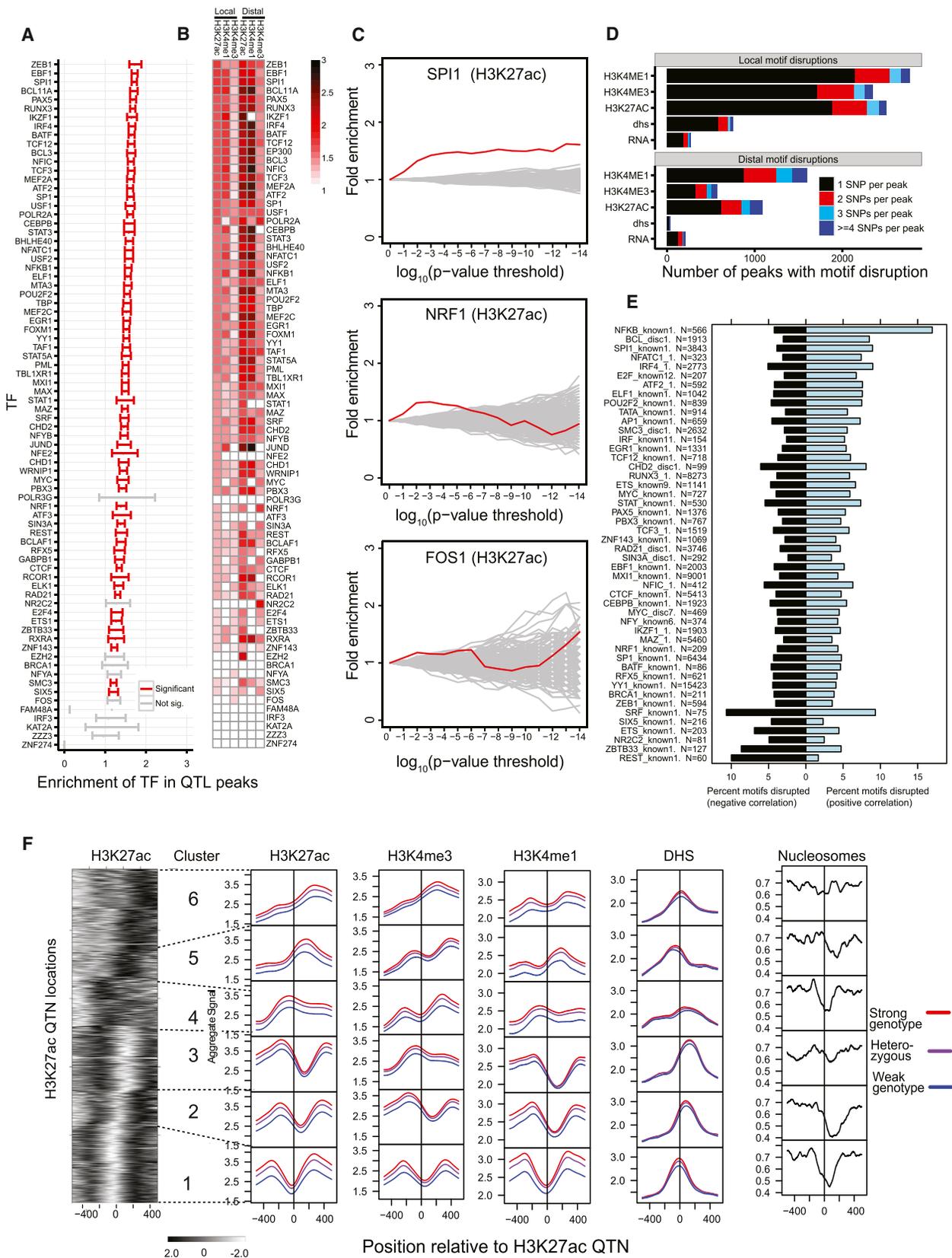
(C) The log odds ratio of observed versus expected state combinations for ChIA-PET-linked peaks. Expected state combinations are calculated by exhaustively counting all possible combinations of states present at ChIA-PET peaks (Fisher's exact test, red = $p < 0.01$).

(D) Enrichment of chromatin interactions in genetic associations (p value $< 10^{-5}$) for each histone mark and RNA. Chromatin interactions were determined by RAD21-ChIA-PET and Hi-C chromatin loops as defined by Rao et al. (2014). The 95% confidence intervals are shown in black ($*p < 0.01$; Fisher's exact test).

(E) Pairs of promoters that are physically linked (ChIA-PET) show correlation of signal for several features. All features except DHS show significantly higher correlation than a control of permuted, distance-matched links or permuted sample labels (Wilcoxon rank sum test).

(F) Same as (E) but for enhancers linked by ChIA-PET. Interacting enhancers are more correlated than any of the permuted, distance-matched data (p value $< 10^{-16}$ for all histone marks and DHS; Wilcoxon rank sum test).

See also Figure S5.



(legend on next page)

with the high frequency of physical interactions within topologically associated domains (TADs), we find that pairs of local-distal hQTLs are enriched within these domains and at looping chromatin contacts. Our results further suggest that specific genetic variants harbored within regulatory elements may concordantly affect local and distal histone modifications through interacting networks of elements. Such concordance could involve a variant in one element influencing another element in physical proximity by recruitment of transcriptional regulators. Alternatively, the process of transcription itself could influence local chromatin (e.g., either directly through polymerase-mediated recruitment of chromatin modifying enzymes, or indirectly by affecting recruitment of histone modifying enzymes by a long non-coding RNA (lncRNA) whose expression is directly affected) or the QTL may influence chromatin state by affecting transcription of enhancer RNAs.

Our results hold special relevance for medical genomics. The robust disease enrichments we observe suggest that long-range histone QTL mapping in diverse cell types will help elucidate the relevant cell type-specific functional elements that give rise to GWAS signal. Furthermore, we propose that hQTL mapping is an important tool for linking GWAS tag SNPs to putative target genes, regulatory elements, and networks bypassing the need to first identify the causal variant (Figure 7). Future investigations of these regulatory networks anchored on GWAS variants may help elucidate the genetic and molecular mechanism underlying diseases and traits.

EXPERIMENTAL PROCEDURES

ChIP-Seq

Cells were cross-linked with formaldehyde at 1% final concentration for 10 min. Chromatin corresponding to 20 million cells was sheared and subjected to immunoprecipitation with antibodies to H3K4me3, H3K4me1, or H3K27ac, respectively. Enriched fragments were subjected to Illumina TruSeq library preparation and paired-end sequencing.

ChIP-Seq, DHS, and RNA-Seq Data Processing

We aligned histone mark ChIP-seq data sets for H3K4me1, H3K4me3 and H3K27ac to personal genomes and subsampled the aligned reads to ensure

consistent sequencing depth. We then called peaks on each individual using MACS (Zhang et al., 2008), merged peaks across all individuals, and extracted signal in each region for each individual. This was followed by quantile normalization, standardization and batch effects removal using PEER (Stegle et al., 2010). We used a similar strategy for the DNase data from (Degner et al., 2012). For quantifying RNA levels from GEUVADIS (Lappalainen et al., 2013), we used Sailfish (Patro et al., 2014), followed by the above normalization scheme on TPM values.

Hi-C

Twenty-five million cells for GM12878 were cross-linked, nuclei-lysed, and chromatin-digested with HindIII. DNA overhangs were biotinylated and proximity ligated under dilute conditions to favor ligation of fragments in three-dimensional proximity. DNA was then sheared, and biotinylated fragments were pulled down with streptavidin beads to enrich for physically interacting sites. Libraries were prepared for Illumina paired-end sequencing and data was processed with the HiCUP pipeline (<http://www.bioinformatics.babraham.ac.uk/projects/hicup/>) to obtain the interaction counts for each pair of restriction fragments. Based on the assumption that two sites that largely interact with the same set of sites are likely to also interact with each other we calculate the proximity between two restriction fragments A and B as the ratio between fragments that interact with both A and B and the number of fragments interacting with only one of them. The threshold of 0.4 for calling an interaction significant was determined based on concordance measures across pseudo replicates (Figure S2; Supplemental Experimental Procedures).

ChIA-PET

Cells for GM12878 were cross-linked and chromatin was prepared for immunoprecipitation with antibodies against RAD21 and H3K4me3, respectively. Immunocomplexes were pulled down with protein-G Dynabeads, marked with biotinylated linkers and subsequently proximity ligated. Biotinylated fragments were pulled down with streptavidin Dynabeads and libraries prepared for Illumina paired-end sequencing. Physical interactions have been called using the Mango pipeline that corrects for known biases in ChIA-PET experiments, such as genomic distance (Phanstiel et al., 2015).

Local hQTL Calling

For identifying the local histone QTLs, we searched for the best correlated SNP within 2 kb of the peak boundaries. The FDR was estimated based on a distribution of empirical p-values obtained from associations from 1,000 sets of peak-wise independent permutations of the sample labels (Figure S1; Supplemental Experimental Procedures).

Figure 6. Transcription Factor Binding Analysis at Local and Distal hQTLs

(A) Overlap enrichment of TF binding in H3K27ac peaks that have a hQTL. For each TF, we plotted enrichment of having an hQTL and being bound by the respective TF. Bars represent the 95% confidence interval. In red are significant enrichments, in gray non-significant ones.

(B) Overlap enrichment of TF binding in histone mark peaks, focusing on peaks with a local hQTL (first three columns) or peaks affecting distal sites (last three columns). The enrichment value is only plotted for TF-histone mark pairs for which the enrichment was significant. Rows are sorted by the fold enrichment for peaks with a local H3K27ac QTL (displayed in A).

(C) Rank enrichment of TF binding in H3K27ac peaks. We plot the fold change enrichment of H3K27ac peaks in TF binding sites at increasing levels of significance for called hQTL peaks (red). The background enrichment was obtained by permuting the p values between peaks (gray).

(D) Number of total peaks that show correlation between a TF motif score and the molecular phenotype. For each molecular phenotype, we computed the correlation between signal and motif score and defined significant motif disruptions using permutations (5% FDR). The number of peaks with at least one significantly correlated TF motif within 2 kb is shown. For each molecular phenotype, peaks were grouped by the number of motif-disrupting SNPs per peak.

(E) Fraction of H3K27ac hQTLs that are significantly correlated with TF motif disruptions. TF motifs show positive and negative correlation with the local histone mark signal and are sorted by the difference between percent positive and negative correlations. The total number of tested SNP-peak pairs across all H3K27ac peaks is annotated next to the TF name. Only TFs with $N \geq 50$ are shown.

(F) The signal surrounding H3K27ac QTNs was extracted and grouped into six clusters (pam clustering, Supplemental Experimental Procedures). The aggregate signals for the six clusters are shown for the high-, heterozygous- and low-genotypes (blue, purple, red) for H3K27ac, H3K4me3, H3K4me1, and DHS. Nucleosome positioning is indicated by MNase signal extracted from the same regions for a single individual (left to right; signal heat maps are shown in Figure S6H). As expected, histone signal coincides with MNase signal/nucleosomes, whereas DNase hypersensitivity coincides with nucleosome-free regions. QTNs show concordant effects on all three histone marks.

See also Figure S6 and Table S3.

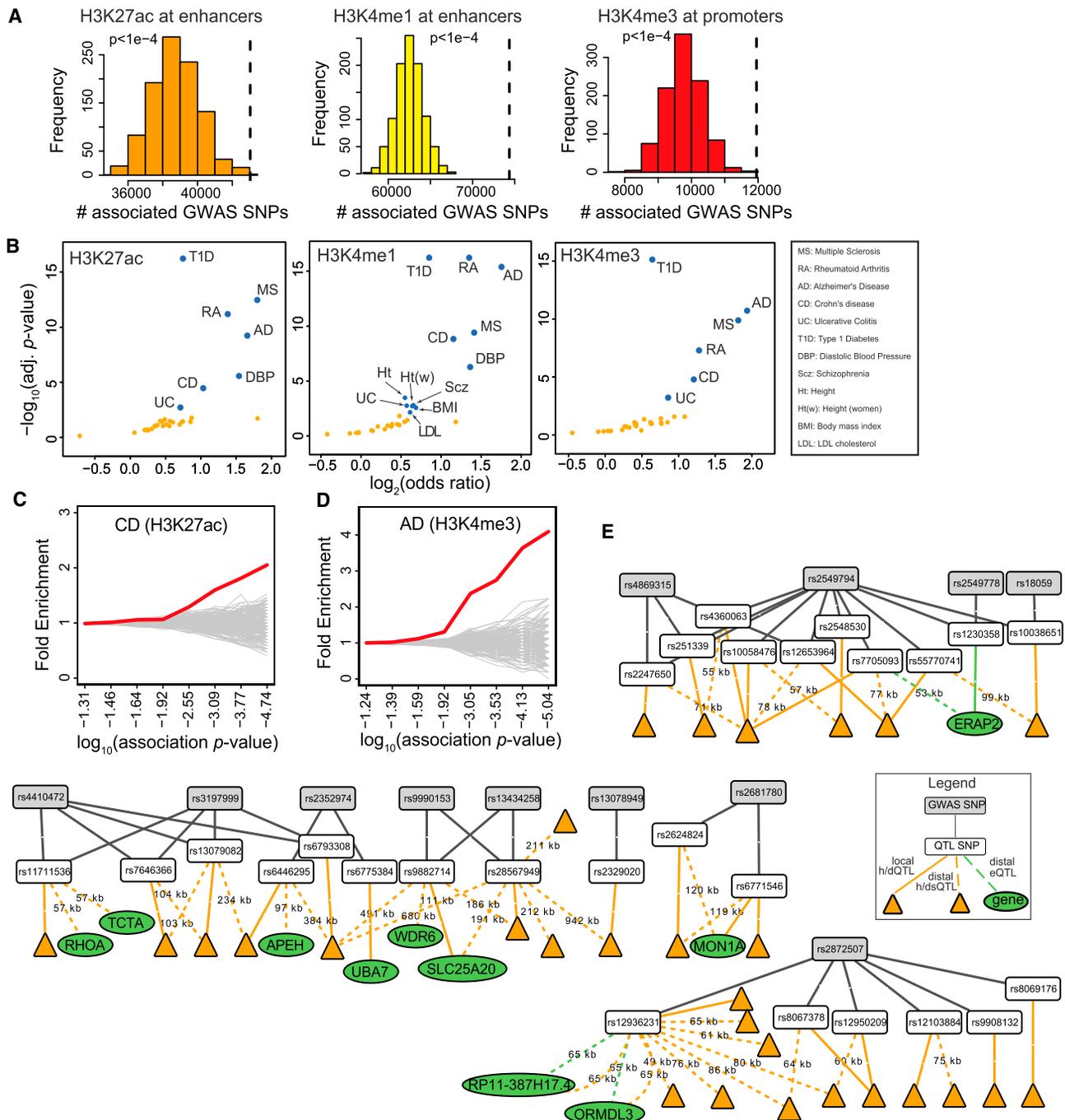


Figure 7. Effect of Regulatory Elements and hQTLs on Phenotypic Diversity

(A) hQTLs are significantly enriched in GWAS SNPs. Dashed lines indicate the value for GWAS SNPs, while the null distributions indicate values by matching on GWAS SNPs for MAF, LD, and distance to TSS. Empirical p values were computed comparing the true to matched null distributions.

(B) Enrichment of hQTLs in GWAS SNPs for specific diseases. We plot the negative log₁₀ adjusted p value by log₂ odds ratio of the Fisher's exact test.

(C and D) Rank enrichment of hQTLs in GWAS SNPs. At each tier of significance for GWAS variants we plot the fold change enrichment in overlap with hQTLs compared to base overlap of all GWAS variants at any significance level (red). We permute the p-values between the GWAS variants to generate background enrichments (gray). Shown here is one example of positive enrichment each for H3K27ac and K3K4me3, respectively.

(E) Regulatory network for Crohn's disease derived from hQTLs and eQTLs intersected with GWAS SNPs. The network consists of (1) GWAS tag SNPs (gray boxes) connected to QTL SNPs (white nodes) through edges representing LD (gray edges), and (2) QTL SNPs connected to affected regulatory elements (orange triangles) or genes (green nodes) through edges representing either local QTLs (solid lines) or distal QTLs (dotted lines). An orange edge represents an hQTL/dsQTL; a green edge represents an eQTL. Distal QTL edges are labeled with the distance between the QTL SNP and the midpoint of the regulatory element.

(legend continued on next page)

Distal hQTL Calling

We only tested SNPs that were associated with a local peak above the 10% FDR threshold for local associations. We then calculated associations for all SNP-peak pairs within 2Mb. To correct for LD and obtain a set of independently associated SNPs for each peak, we applied the following algorithm: (1) identify and keep the most significant SNP (regardless of distance) for each peak, (2) remove any SNP in LD ($r^2 > 0.2$), repeat (1) and (2) until no SNP was left (Figure S3D; Supplemental Experimental Procedures). We then split the set of SNP-peak pairs into Hi-C-interacting and non-interacting pairs and calculated the FDR independently for each set using the a permutation-based empirical p value distribution (1,000 sets of per-peak independent sample label permutations).

Enrichment in Topological Domains

To assess the enrichment of distal QTLs within TADs we (1) calculated the overlap relative to a set of shuffled domains, and (2) compared the fraction of same-domain QTL pairs to random sets of distance-matched SNP-peak pairs (only SNPs with a local QTL were used in the construction of this control set) (Figures 3F and 3G).

TF Motif Analysis

We tested enrichment of TF binding sites in the peaks that have QTL using a Fisher's exact test. Then, we checked whether TF motif disruption (computed as TF motif PWM scores computed at each SNP per individual) correlates with the signal at histone marks and DHS, or with RNA expression, using Spearman correlation. We assessed significance by permutation testing with an FDR of 5%.

GWAS Analysis

We overlapped hQTL SNPs for H3K27ac, H3K4me, and H3K4me3 with GWAS SNPs with $p < 1e-5$ for over 150 diseases. We compared the overlap to two negative sets: the set of GWAS SNPs with $p > 1e-5$ and a set of SNPs matched for MAF, LD, and distance to TSS. We assessed difference in enrichment with a one-sided Fisher's exact test. We first thinned GWAS SNPs to keep only the most significant variant of those variants in EUR LD with $r^2 > 0.8$ and then considered an overlap with any SNP in EUR LD $r^2 > 0.8$ with the remaining GWAS SNPs.

ACCESSION NUMBERS

The accession number for the data sets reported in this paper is Gene Expression Omnibus (GEO): GSE62742. The data and analysis files are at <http://chromovar3d.stanford.edu/> and <http://www.zaugg.embl.de/data-and-tools/distal-chromatin-qtls/>, code at <https://github.com/kundajelab/chromovar3d>.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and four tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.07.048>.

AUTHOR CONTRIBUTIONS

F.G., J.B.Z., M.K., and O.U. conceptualized the study. F.G., J.B.Z., M.K., O.U., D.S., A.R.M., P.G., R.S., D.H.P., A.P., N.H., and G.E. developed methodology. F.G., M.K., and D.S. performed investigation. J.B.Z., F.G., O.U., M.K., A.R.M., P.G., R.S., and D.H.P. performed formal analysis. F.G., J.B.Z., M.K., O.U., A.R.M., P.G., R.S., A.K., and M.S. wrote the original draft of the manuscript.

F.G., J.B.Z., M.K., O.U., A.R.M., P.G., R.S., D.H.P., D.S., A.K., and M.S. reviewed and edited the manuscript. M.S. provided funding. M.S. and A.K. provided resources. M.S., A.K., L.M.S., C.D.B., J.K.P., and W.H. supervised the study.

ACKNOWLEDGMENTS

Our work was supported by grants from the Vera Moulton Wall Center for Pulmonary Vascular Disease (F.G., M.K.), Swiss National Foundation (J.B.Z.), NIH Medical Scientist Training Program Training Grant T32GM007205 (M.K.), Gabilan Stanford Graduate Fellowship (O.U.), NIH/NHGRI T32 HG000044 and the Genentech Graduate Fellowship (D.V.S), NIH Genetics and Developmental Biology Training Program T32 GM007790 (A.R.M), the Stanford Biomedical Informatics Training Grant from the National Library of Medicine (LM-07033) (P.G.), R.S. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (DRG-2187-14), D.H.P. is a Damon Runyon fellow supported by the Damon Runyon Cancer Research Foundation (DRG-2122-12), EC FP7 project "RADIANT" (A.P., W.H.), NIH and from the European Research Council (L.M.S), Alfred Sloan Foundation and NIH R01ES025009 and NIH U41-HG007000-02S1 (subcontract) (A.K.), NIH 5U54HG006996-03 and 5U01HL107393-04, and Genetics Department, Stanford University (M.S.). C.D.B. is a founder of Identify Genomics and a science advisory board member for AncestryDNA, InVita, Personalis and Identify Genomics; L.M.S. is a founder and member of the science advisory board for Sophia Genetics; M.S. is a founder and member of the science advisory board of Personalis and a science advisory board member of Genapsys and AxiomX. We would like to thank Prof. Stephen Montgomery and members of the M.S. lab for critical comments on the manuscript. We would also like to thank Gerald Quon and Abhishek Sarkar for sharing GWAS summary statistics and SCGPM and SCG3 administrators for compute support and supplementary website hosting.

Received: October 16, 2014

Revised: March 5, 2015

Accepted: July 21, 2015

Published: August 20, 2015

REFERENCES

- Bourgon, R., Gentleman, R., and Huber, W. (2010). Independent filtering increases detection power for high-throughput experiments. *Proc. Natl. Acad. Sci. USA* 107, 9546–9551.
- Degner, J.F., Pai, A.A., Pique-Regi, R., Veyrieras, J.B., Gaffney, D.J., Pickrell, J.K., De Leon, S., Michelini, K., Lewellen, N., Crawford, G.E., et al. (2012). DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* 482, 390–394.
- Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380.
- Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Fritze, S., Harrow, J., Kaul, R., et al.; ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74.
- Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shores, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473, 43–49.

Regulatory elements are defined as the merged peaks from H3K4me1, H3K4me3, H3K27ac, and DHS. If the affected peak in a regulatory element overlaps the transcription start site of a gene, we label the regulatory element with the gene's name. If multiple QTL SNPs are associated with the same local regulatory element, we label the QTL SNP with the reference SNP cluster identifier (rsID) of the most significant SNP; if multiple QTL SNPs are associated with the same distal regulatory element, then for every combination of local-distal association we pick the best-correlated SNP for the local peak. (Here, we only show the parts of the network that contain a gene. For the full network see Figure S7.)

See also Figure S7 and Table S4.

- Franke, A., McGovern, D.P., Barrett, J.C., Wang, K., Radford-Smith, G.L., Ahmad, T., Lees, C.W., Balschun, T., Lee, J., Roberts, R., et al. (2010). Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat. Genet.* *42*, 1118–1125.
- Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., et al. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* *462*, 58–64.
- Gjoneska, E., Pfenning, A.R., Mathys, H., Quon, G., Kundaje, A., Tsai, L.H., and Kellis, M. (2015). Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature* *518*, 365–369.
- Harismendy, O., Notani, D., Song, X., Rahim, N.G., Tanasa, B., Heintzman, N., Ren, B., Fu, X.D., Topol, E.J., Rosenfeld, M.G., and Frazer, K.A. (2011). 9p21 DNA variants associated with coronary artery disease impair interferon- γ signalling response. *Nature* *470*, 264–268.
- Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* *22*, 1760–1774.
- Hindorf, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* *106*, 9362–9367.
- Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* *503*, 290–294.
- Karczewski, K.J., Dudley, J.T., Kukurba, K.R., Chen, R., Butte, A.J., Montgomery, S.B., and Snyder, M. (2013). Systematic functional regulatory assessment of disease-associated variants. *Proc. Natl. Acad. Sci. USA* *110*, 9607–9612.
- Kasowski, M., Kyriazopoulou-Panagiotopoulou, S., Grubert, F., Zaugg, J.B., Kundaje, A., Liu, Y., Boyle, A.P., Zhang, Q.C., Zakharia, F., Spacek, D.V., et al. (2013). Extensive variation in chromatin states across humans. *Science* *342*, 750–752.
- Kheradpour, P., and Kellis, M. (2014). Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res.* *42*, 2976–2987.
- Kilpinen, H., Waszak, S.M., Gschwind, A.R., Raghav, S.K., Witwicki, R.M., Orjoli, A., Migliavacca, E., Wiederkehr, M., Gutierrez-Arcelus, M., Panousis, N.I., et al. (2013). Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science* *342*, 744–747.
- Kleinjan, D.A., and van Heyningen, V. (2005). Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am. J. Hum. Genet.* *76*, 8–32.
- Koren, A., Polak, P., Nemes, J., Michaelson, J.J., Sebat, J., Sunyaev, S.R., and McCarroll, S.A. (2012). Differential relationship of DNA replication timing to different forms of human mutation and variation. *Am. J. Hum. Genet.* *91*, 1033–1040.
- Kundaje, A., Kyriazopoulou-Panagiotopoulou, S., Libbrecht, M., Smith, C.L., Raha, D., Winters, E.E., Johnson, S.M., Snyder, M., Batzoglu, S., and Sidow, A. (2012). Ubiquitous heterogeneity and asymmetry of the chromatin environment at regulatory elements. *Genome Res.* *22*, 1735–1747.
- Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., et al.; European Alzheimer's Disease Initiative (EADI); Genetic and Environmental Risk in Alzheimer's Disease; Alzheimer's Disease Genetic Consortium; Cohorts for Heart and Aging Research in Genomic Epidemiology (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* *45*, 1452–1458.
- Lappalainen, T., Sammeth, M., Friedländer, M.R., 't Hoen, P.A., Monlong, J., Rivas, M.A., González-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al.; Geuvadis Consortium (2013). Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* *501*, 506–511.
- Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* *148*, 84–98.
- Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragozcy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* *326*, 289–293.
- Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science* *337*, 1190–1195.
- McVicker, G., van de Geijn, B., Degner, J.F., Cain, C.E., Banovich, N.E., Raj, A., Lewellen, N., Myrthil, M., Gilad, Y., and Pritchard, J.K. (2013). Identification of genetic variants that affect histone modifications in human cells. *Science* *342*, 747–749.
- Nicolae, D.L., Gamazon, E., Zhang, W., Duan, S., Dolan, M.E., and Cox, N.J. (2010). Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* *6*, e1000888.
- Nobrega, M.A., Ovcharenko, I., Afzal, V., and Rubin, E.M. (2003). Scanning human gene deserts for long-range enhancers. *Science* *302*, 413.
- Patro, R., Mount, S.M., and Kingsford, C. (2014). Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat. Biotechnol.* *32*, 462–464.
- Phanstiel, D.H., Boyle, A.P., Heidari, N., and Snyder, M.P. (2015). Mango: a bias-correcting ChIA-PET analysis pipeline. *Bioinformatics*, btv336.
- Pomerantz, M.M., Ahmadiyeh, N., Jia, L., Herman, P., Verzi, M.P., Doddapaneni, H., Beckwith, C.A., Chan, J.A., Hills, A., Davis, M., et al. (2009). The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat. Genet.* *41*, 882–884.
- Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* *159*, 1665–1680.
- Rivas, M.A., Beaudoin, M., Gardet, A., Stevens, C., Sharma, Y., Zhang, C.K., Boucher, G., Ripke, S., Ellinghaus, D., Burtt, N., et al.; National Institute of Diabetes and Digestive Kidney Diseases Inflammatory Bowel Disease Genetics Consortium (NIDDK IBDGC); United Kingdom Inflammatory Bowel Disease Genetics Consortium; International Inflammatory Bowel Disease Genetics Consortium (2011). Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat. Genet.* *43*, 1066–1073.
- Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature* *489*, 109–113.
- Schaub, M.A., Boyle, A.P., Kundaje, A., Batzoglu, S., and Snyder, M. (2012). Linking disease associations with regulatory information in the human genome. *Genome Res.* *22*, 1748–1759.
- Smemo, S., Tena, J.J., Kim, K.H., Gamazon, E.R., Sakabe, N.J., Gómez-Marín, C., Aneas, I., Credidio, F.L., Sobreira, D.R., Wasserman, N.F., et al. (2014). Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* *507*, 371–375.
- Stegle, O., Parts, L., Durbin, R., and Winn, J. (2010). A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLoS Comput. Biol.* *6*, e1000770.
- Van Limbergen, J., Wilson, D.C., and Satsangi, J. (2009). The genetics of Crohn's disease. *Annu. Rev. Genomics Hum. Genet.* *10*, 89–116.
- Visel, A., Rubin, E.M., and Pennacchio, L.A. (2009). Genomic views of distant-acting enhancers. *Nature* *461*, 199–205.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* *9*, R137.
- Zuk, O., Schaffner, S.F., Samocha, K., Do, R., Hechter, E., Kathiresan, S., Daly, M.J., Neale, B.M., Sunyaev, S.R., and Lander, E.S. (2014). Searching for missing heritability: designing rare variant association studies. *Proc. Natl. Acad. Sci. USA* *111*, E455–E464.