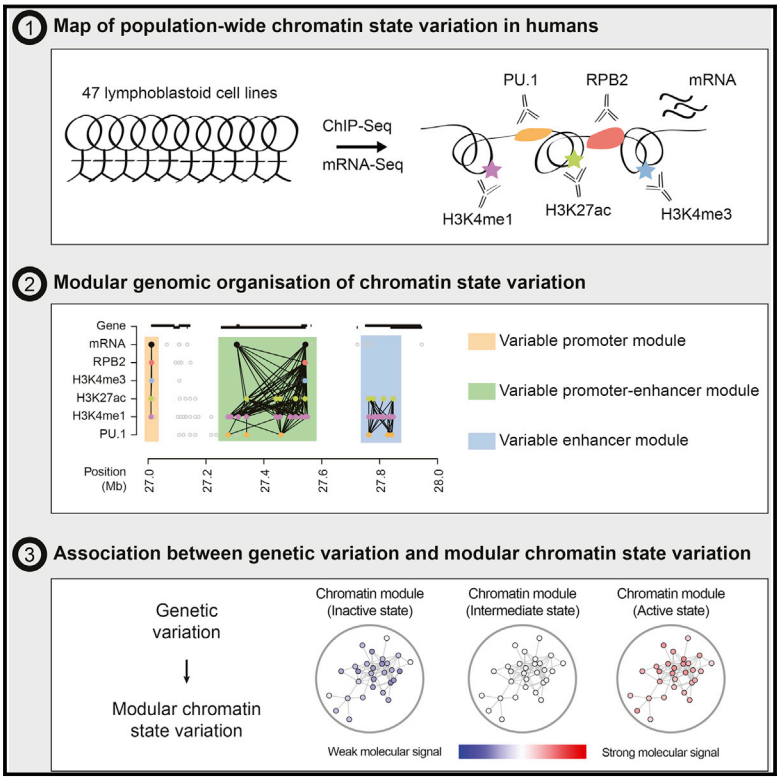
Article

Cell

Population Variation and Genetic Control of Modular Chromatin Architecture in Humans

Graphical Abstract



Highlights

- Modules of correlated molecular phenotypes represent inter-individual chromatin variation
- Variable chromatin modules (VCMs) are embedded within chromosomal contact domains
- VCMs are orchestrated by *cis*-acting genetic variation
- VCMs rationalize chromatin state changes that are independent of local DNA sequence

Authors

Sebastian M. Waszak, Olivier Delaneau, Andreas R. Gschwind, ..., Alexandre Reymond, Bart Deplancke, Emmanouil T. Dermitzakis

Correspondence

bart.deplancke@epfl.ch (B.D.), emmanouil.dermitzakis@unige.ch (E.T.D.)

In Brief

Spatially defined chromosome regions, termed variable chromatin modules, exhibit coordinated chromatin state changes across *cis*-regulatory elements. Within these modules, genetic changes distal from the regulatory element itself can induce variation in chromatin patterns between individuals.

Waszak et al., 2015, Cell *162*, 1039–1050
August 27, 2015
http://dx.doi.org/10.1016/j.cell.2015.08.001

CellPress

Cell

Population Variation and Genetic Control of Modular Chromatin Architecture in Humans

Sebastian M. Waszak,[1,2,7,9] Olivier Delaneau,[2,3,4,7] Andreas R. Gschwind,[2,5] Helena Kilpinen,[2,3,4] Sunil K. Raghav,[1] Robert M. Witwicki,[5] Andrea Orioli,[5] Michael Wiederkehr,[5] Nikolaos I. Panousis,[2,3,4] Alisa Yurovsky,[2,3,4] Luciana Romano-Palumbo,[3] Alexandra Planchon,[3] Deborah Bielser,[3] Ismael Padioleau,[2,3,4] Gilles Udin,[1] Sarah Thurnheer,[6] David Hacker,[6] Nouria Hernandez,[5] Alexandre Reymond,[5,8] Bart Deplancke,[1,2,8,*] and Emmanouil T. Dermitzakis[2,3,4,8,*]

[1]Institute of Bioengineering, School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne 1015, Switzerland
[2]Swiss Institute of Bioinformatics, Lausanne 1015, Switzerland
[3]Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva 1211, Switzerland
[4]Institute of Genetics and Genomics in Geneva, University of Geneva, Geneva 1211, Switzerland
[5]Center for Integrative Genomics, Faculty of Biology and Medicine, University of Lausanne, Lausanne 1015, Switzerland
[6]Protein Expression Core Facility, School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne 1015, Switzerland
[7]Co-first author
[8]Co-senior author
[9]Present address: Genome Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg 69117, Germany
*Correspondence: bart.deplancke@epfl.ch (B.D.), emmanouil.dermitzakis@unige.ch (E.T.D.)
http://dx.doi.org/10.1016/j.cell.2015.08.001

SUMMARY

Chromatin state variation at gene regulatory elements is abundant across individuals, yet we understand little about the genetic basis of this variability. Here, we profiled several histone modifications, the transcription factor (TF) PU.1, RNA polymerase II, and gene expression in lymphoblastoid cell lines from 47 whole-genome sequenced individuals. We observed that distinct *cis*-regulatory elements exhibit coordinated chromatin variation across individuals in the form of variable chromatin modules (VCMs) at sub-Mb scale. VCMs were associated with thousands of genes and preferentially cluster within chromosomal contact domains. We mapped strong proximal and weak, yet more ubiquitous, distal-acting chromatin quantitative trait loci (cQTL) that frequently explain this variation. cQTLs were associated with molecular activity at clusters of *cis*-regulatory elements and mapped preferentially within TF-bound regions. We propose that local, sequence-independent chromatin variation emerges as a result of genetic perturbations in cooperative interactions between *cis*-regulatory elements that are located within the same genomic domain.

INTRODUCTION

Understanding the genetic contribution and molecular paths toward complex traits is one of the key outstanding challenges in biology. Genome-wide studies revealed that most common disease-associated genetic variants fall into gene regulatory sequences (Manolio, 2010; Maurano et al., 2012; Nica et al., 2010; Nicolae et al., 2010) and affect transcriptional programs in disease-implicated cell types (Fairfax et al., 2012; Grundberg et al., 2012). Evolutionary studies have further uncovered several instances of gene regulatory changes that are causally implicated in complex phenotypes (Wray, 2007). These changes are thought to originate mostly from variation in TF-DNA interactions, which are well known to mediate the spatiotemporal control of gene expression programs (Spitz and Furlong, 2012). Understanding the extent of, and the mechanisms underlying, TF DNA binding variation is therefore key to elucidate the molecular determinants of complex phenotypes. Small-scale population- and family-based studies have shown that 5%–25% of TF-DNA binding events exhibit intra- and inter-individual binding variation (Kasowski et al., 2010, 2013; Kilpinen et al., 2013; McVicker et al., 2013; Reddy et al., 2012). These studies, as well as those examining TF-DNA binding divergence among mammalian species (reviewed in Villar et al. [2014]) showed that only a minority of this variation could be attributed to genetic differences within TF-bound sequences.

So far, few mechanisms have been proposed to clarify this phenomenon, and these are mostly centered on changes in either the local DNA structure or in collaborative interactions between co-bound TFs at *cis*-regulatory elements (Albert and Kruglyak, 2015; Heinz et al., 2013; Karczewski et al., 2011; Kilpinen et al., 2013; Stefflova et al., 2013). Recently, others and we have observed that several chromatin state components exhibit a strong degree of coordinated allelic variation that extends over several thousands of base pairs (Kilpinen et al., 2013; McVicker et al., 2013). This observation suggests that variation in TF-DNA binding might be conditioned on the state of other *cis*-regulatory elements, but a general description of this effect has so far been hampered due to sparseness of allelic markers.

Here, we measured ChIP-seq-based, population-level histone modification (HM) and TF enrichment patterns. Specifically, we mapped the regulatory TF PU.1, the second largest subunit of RNA polymerase II (RPB2), and three well-studied HMs often

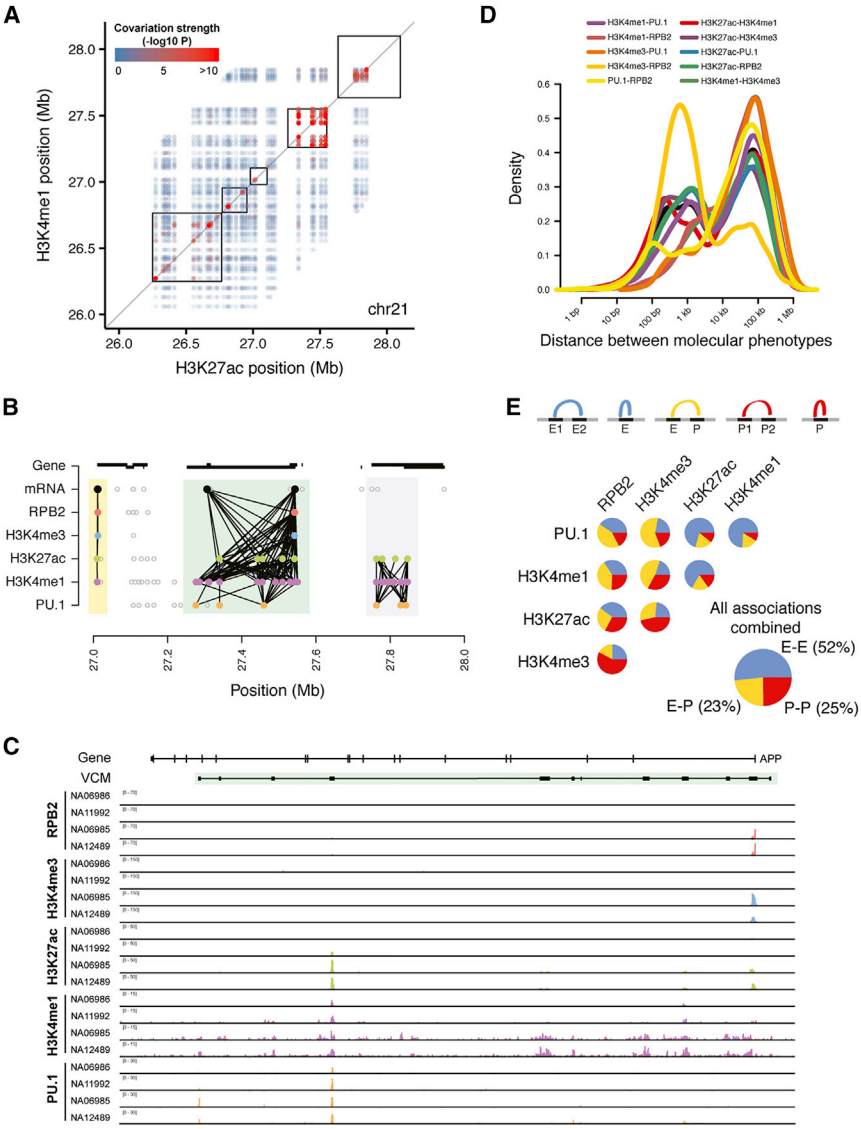


Figure 1. Genome-wide Associations among Molecular Phenotypes

(A) Inter-individual association between the read depth at H3K27ac and H3K4me1 ChIP-seq peaks on chromosome 21 (26,000,000–28,000,000). The pairwise association strength (Pearson's p value) is color coded and ranges from blue (p = 1) to red (p < 1e-10). Chromosomal contact domains (Rao et al., 2014) are shown with black boxes. See Figure S1H for molecular associations in this region based on other marks.

(B) Significant associations between molecular phenotypes in a 1 Mb window on chr21 (27,000,000–28,000,000). Circles indicate variable (filled) or non-variable (open) enrichment of molecular marks (i.e., ChIP-seq peaks or gene expression). Lines connecting filled circles represent significant associations between molecular phenotypes (FDR 0.1%).

(C) Selected individuals with either low (NA06986 and NA11992) or high (NA06985 and NA12489) enrichment of molecular marks around the APP gene locus.

(D) Distance distribution between coordinated molecular phenotypes.

(E) Annotation of *cis*-regulatory elements with coordinated enrichment of molecular marks into putative enhancers (E) and promoters (P).

See also Figures S1, S2, and S3.

observed at enhancers and promoters (H3K4me1, H3K4me3, and H3K27ac) in lymphoblastoid cell lines (LCLs) derived from 47 unrelated European individuals whose genomes were sequenced in the frame of the 1000 Genomes Project (Abecasis et al., 2010). In addition, we profiled gene expression using mRNA sequencing in 46 LCLs. Our results provide unique insights into the mechanisms underlying variation in molecular activity at *cis*-regulatory elements, revealing that most of this variation results from alterations in the modular organization of the human genome.

RESULTS

Population-Level Variation in Molecular Activity at *cis*-Regulatory Elements

To assess the extent of quantitative coordination in inter-individual chromatin variation at putative *cis*-regulatory elements, we performed an association analysis between molecular phenotypes, with "molecular phenotype" being here defined as the normalized and covariate-corrected read depth of a histone-modified and TF-bound region, respectively. Specifically, we estimated the correlation levels between all distinct TF-TF, HM-HM, and TF-HM combinations in 1 Mb *cis* windows (Figure 1A). We tested a total of 29 million associations between any two molecular phenotypes and estimated for each association pair the enrichment of low p values using π_1 statistics (Storey and Tibshirani, 2003). Estimates of π_1 ranged from 2.5% for PU.1-H3K4me3 to 11% for H3K4me1-H3K27ac (Figure S1A), indicating extensive quantitative coordination in molecular activity levels between/at *cis*-regulatory elements. Moreover, molecular coordination decayed quickly with increasing genomic distance and was 20-fold more enriched between proximal *cis*-regulatory elements (<10 kb) than between any two *cis*-regulatory elements that were separated by 500 kb or more (Figure S1B).

Overall, we detected 79,411 statistically significant, mostly positive (>99%) associations (at genome-wide correction) across all molecular association tests (Pearson r_{mean} = 0.70, FDR 0.1%) (Figures 1B, 1C, S1C, and S1D), involving on average 20% of all studied TF-bound/HM-enriched regions (Figure S1E). The histone mark H3K27ac exhibited the highest number and proportion of significant associations with other phenotypes (Figures S1E and S1F), suggesting that this molecular phenotype is most

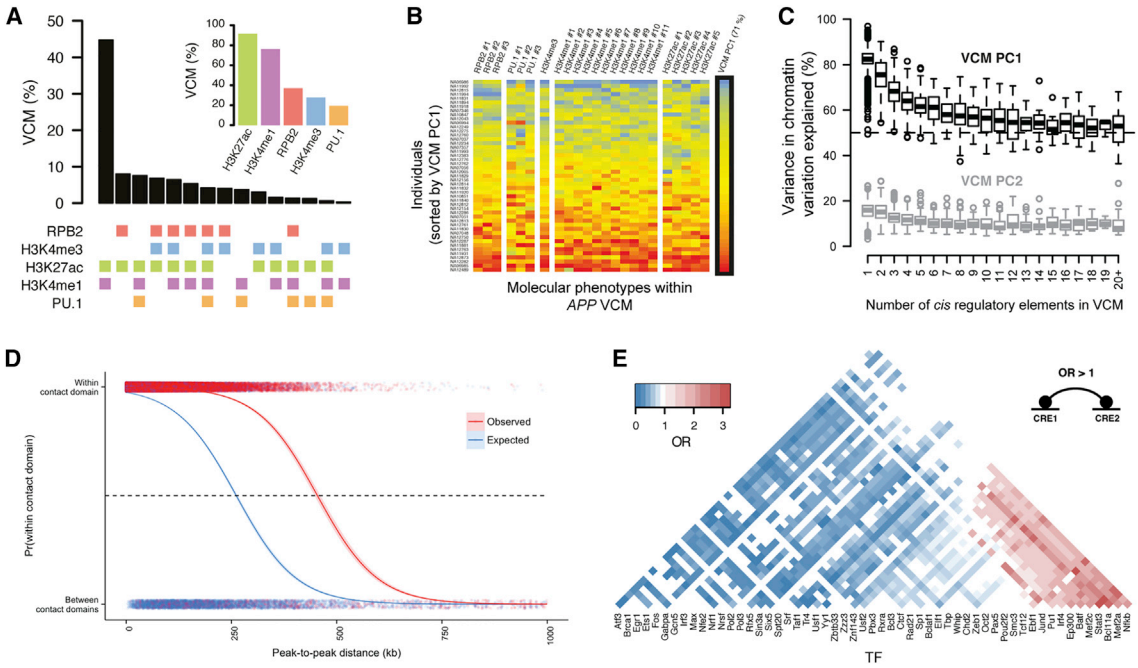


Figure 2. Variable Chromatin Modules

(A) Molecular mark composition of variable chromatin modules (VCMs). Black bars (top) indicate the percentage of VCMs with specific combinations of molecular marks (bottom). Inset shows the percentage of VCMs with a specific molecular mark.

(B) Coordination of molecular activity within VCMs. The heatmap illustrates for 47 individuals (rows) the normalized signal of molecular phenotypes (columns) that belong to the VCM spanning the APP gene locus (as shown in Figures 1B and 1C). (Right column) The first principal component summarizes the majority (71%) of molecular variation within this VCM.

(C) Percentage of molecular variation within VCMs that is explained by the first and second principal components. VCMs were divided according to the number of *cis*-regulatory elements (domains). VCMs with $\geq$20 domains were grouped.

(D) Enrichment of covariable *cis*-regulatory elements within chromosomal contact domains (Rao et al., 2014). (Red) Covariable *cis*-regulatory elements; (blue) random pairs of *cis*-regulatory elements. The probability indicates whether two covariable *cis*-regulatory elements are embedded within the same contact domain as opposed to two distinct contact domains.

(E) Co-associations of TF-TF pairs at non-overlapping, covariable *cis*-regulatory elements. Positive and negative odds ratios indicate significant enrichment/depletion of TF-TF pairs ($p < 0.05$ after Bonferroni correction).

See also Figures S2 and S3.

sensitive to coordinated chromatin state perturbations. As expected, the TFs PU.1 and RPB2 were preferentially associated with enhancer- (H3K27ac/H3K4me1 for PU.1) and promoter-marking HMs (H3K27ac/H3K4me3 for RPB2), respectively (Figure S1G). Except for RPB2-H3K4me3, the majority of all molecular associations were identified between non-overlapping *cis*-regulatory elements (Figure S2A), which exhibit a log-normal distance distribution that preferentially centered around 45 kb (95% confidence interval [CI]: 7–308 kb) (Figures 1D and S2B). The molecular association strength between covariable *cis*-regulatory elements decayed significantly with increasing distance ($\rho = -0.19$, $p < 2.2e\text{-}16$, Figure S2C). Overall, 25% of all molecular associations were found between promoters and enhancers (>5 kb from transcription start site [TSS]), 25% within or between promoters, and 50% within or between putative enhancers (Figure 1E). These results suggest extensive molecular coupling between *cis*-regulatory elements and a strong degree of chromatin variation at enhancer-like regions.

The previous results indicate that chromatin state variation might reflect high-order genomic interactions. Using simple graph-based methods, we could map individual molecular associations into 14,559 distinct "variable chromatin modules (VCMs)" that are composed of 25,417 distinct *cis*-regulatory elements (see Figures 1B, 1C, and S3A–S3C for examples). The median size of a single VCM was 4.2 kb, and all molecular phenotypes contained within VCMs together covered 5% (161 Mb) of the human genome. Although only 25% of VCMs were composed of multiple *cis*-regulatory elements (Figure S3D), these "multi-VCMs" captured (1) the vast majority (78%) of molecular associations (Figure S3E), (2) were more likely to contain promoter- and enhancer-marking chromatin marks (Figure S3F), and (3) covered more DNA sequence (median size: 70 kb; Figure S3G).

The majority of VCMs (56%) were exclusively composed of enhancer-marking signals (i.e., H3K4me1-PU.1, H3K4me1-H3K27ac, and H3K4me1-H3K27ac-PU.1) (Figure 2A), indicating that putative enhancers constitute the largest part of the variable epigenome in a single human population, consistent with comparative epigenomic studies across mammalian species (Villar et al., 2015).

To examine the extent of molecular coordination within VCMs, we tested whether the activity state of a VCM can be represented by a single quantitative phenotype, rather than by individual

molecular phenotypes that define a VCM. We applied principal component (PC) analysis and extracted the first and second PC for each VCM (Figure 2B). We found that the first PC already explains, on average, 79% of the variability that is observed between molecular phenotypes of the same VCM (Figure 2C), suggesting that molecular activity is strongly coordinated within VCMs.

This high degree of molecular coordination within VCMs implies a higher-order chromatin organization, consistent with the now well-accepted notion that mammalian genomes are spatially arranged in distinct chromosomal contact domains (Dixon et al., 2012; Rao et al., 2014). To test this hypothesis, we analyzed published, high-resolution, and genome-wide chromatin conformation data from a human lymphoblastoid cell line (Rao et al., 2014) and found that *cis*-regulatory elements with coordinated chromatin state variation were more preferentially embedded within the same chromosomal contact domain (odds ratio = 14.9, p = 2.2e-16, logistic regression) (Figure 2D; see Figures 1A and S1H-S1I for examples). We also observed that *cis*-regulatory elements of the same VCM exhibited more frequently allelic chromatin biases along the same haplotype (OR = 1.3, p = 4.9e-5, logistic regression), further indicating that VCM define a regulatory unit. Moreover, analysis of genome-wide TF-DNA binding data of the architectural proteins CTCF and cohesin (RAD21/SMC3) (Ong and Corces, 2014) revealed a significant enrichment at *cis*-regulatory elements that participate in long-range (>300–500 kb) molecular associations (Figures S2D–S2F). Together, these results support our hypothesis that VCMs represent a fine-grained, modular architecture of the variable human epigenome.

Next, we aimed to elucidate mechanisms that may be responsible for the emergence of VCMs. Here, we hypothesized that modular chromatin state dynamics may not only be driven by short-range cooperative TF-TF interactions, as shown earlier (Karczewski et al., 2011; Kasowski et al., 2010; Kilpinen et al., 2013; Zheng et al., 2010), but also by interactions that act over long genomic distances and across *cis*-regulatory elements. To test this hypothesis, we investigated whether particular TF-TF pairs exhibited preferential enrichments at pairs of *cis*-regulatory elements that are part of the same VCM using experimentally defined TF-DNA binding data (ENCODE Project Consortium, 2012). This analysis revealed 204 putative cooperative TF-TF pairs that are preferentially enriched at VCM-defined *cis*-regulatory elements (OR = 1.1-3.2; $p < 0.05$ after Bonferroni correction; Fisher's exact test) (Figure 2E). For example, NFKB emerged as the most cooperative TF among all tested factors and was preferentially associated with well-known immunity-associated TFs (e.g., STAT3, BCL11A, BATF, and PU.1). Thus, our results suggest that modular chromatin dynamics occur within spatially organized domains of the genome and are likely, in part, mediated by long-range cooperative interactions between TFs that determine the molecular identity of a lymphoblastoid cell (Zhou et al., 2015).

Chromatin Variation Reflects Inter-Individual Variation in Gene Expression

To assess the functional impact of inter-individual chromatin state variation, we analyzed associations in *cis* between molecular phenotypes at *cis*-regulatory elements and gene expression (TSS ± 1 Mb). This analysis resulted in significant associations for 4,568 (22%) genes at a FDR of 0.1% (Figure S3H and see Figures 3A, S3I, and S3J for examples). The vast majority (99%) of chromatin-gene associations were positive (i.e., higher gene expression levels correlated with stronger chromatin signals) (Figure S3K), explained about half of the variation in gene expression (Figure S3L), and correlated independently with multiple molecular events at *cis*-regulatory elements. Two-thirds of all gene-associated *cis*-regulatory elements mapped outside of promoters (TSS ± 2.5 kb) and thus likely pinpoint to putative enhancer-gene interactions (Figures 3B and 3C). We further measured allelic expression effects within individuals and observed that, consistent with coordinated allelic chromatin signals, they are more concordant with allelic chromatin states at gene-associated regions than at random regions (OR = 1.9, p = 2e-10, logistic regression). Together, these results provide genome-wide evidence that population-level variation in chromatin states has functional consequences and that it is a potential approach to identify the gene targets of putative *cis*-regulatory elements.

We also observed that VCM states (as defined by the first PC) were associated with 3,580 genes in *cis* (TSS ± 1 Mb; FDR 0.1%). This analysis has further allowed us to uncover that only 5% of "enhancer VCMs" (H3K27ac-H3K4me1-PU.1) varied along with nearby genes despite representing the most abundant class of VCMs. In strong contrast, variable promoter (H3K27ac-H3K4me3-RPB2) and promoter-enhancer (H3K27ac-H3K4me3-H3K4me1-RPB2-PU.1) VCMs correlated with gene expression in up to 80% of the cases (Figure 3D). Moreover, 23% of all gene-associated VCMs correlated with the expression levels of multiple genes (Figure S3M), suggesting that these VCMs contain *cis*-regulatory elements that are potentially shared across genes. We also found that VCMs with several *cis*-regulatory elements were more likely to reflect variable gene expression (Spearman's $\rho = 0.91$, p = 1.8e-8) (Figure 3E), suggesting that both the type (promoter/enhancer) and number of variable *cis*-regulatory elements are key determinants underlying the transcriptional state change of a gene.

We next assessed whether VCMs were located nearby specific sets of genes and found that VCMs embedding several *cis*-regulatory elements were highly enriched in immunity-related processes and pathways (Tables S2A and S2B), consistent with the biological nature of lymphoblastoid cells. Functional analysis of chromatin-associated genes further supported a strong enrichment of VCMs in immunity-related processes (Table S2C).

Genetic Control of Chromatin State and Gene Expression Variation

To identify potential mechanisms that explain variation in TF-DNA binding, HMs, VCM states, and gene expression, we mapped quantitative trait loci (QTLs) for all studied molecular phenotypes independently in a 500 kb *cis*-window around the center of a candidate *cis*-regulatory element (or TSS). We detected between 315 and 1,432 significant chromatin QTLs (cQTLs, i.e., tfQTLs and hmQTLs) and eQTLs at 10% FDR. This corresponds from 1.1% (H3K4me1) to 2.9% (mRNA) of the studied regions and explained ~40% of their variability (Figures

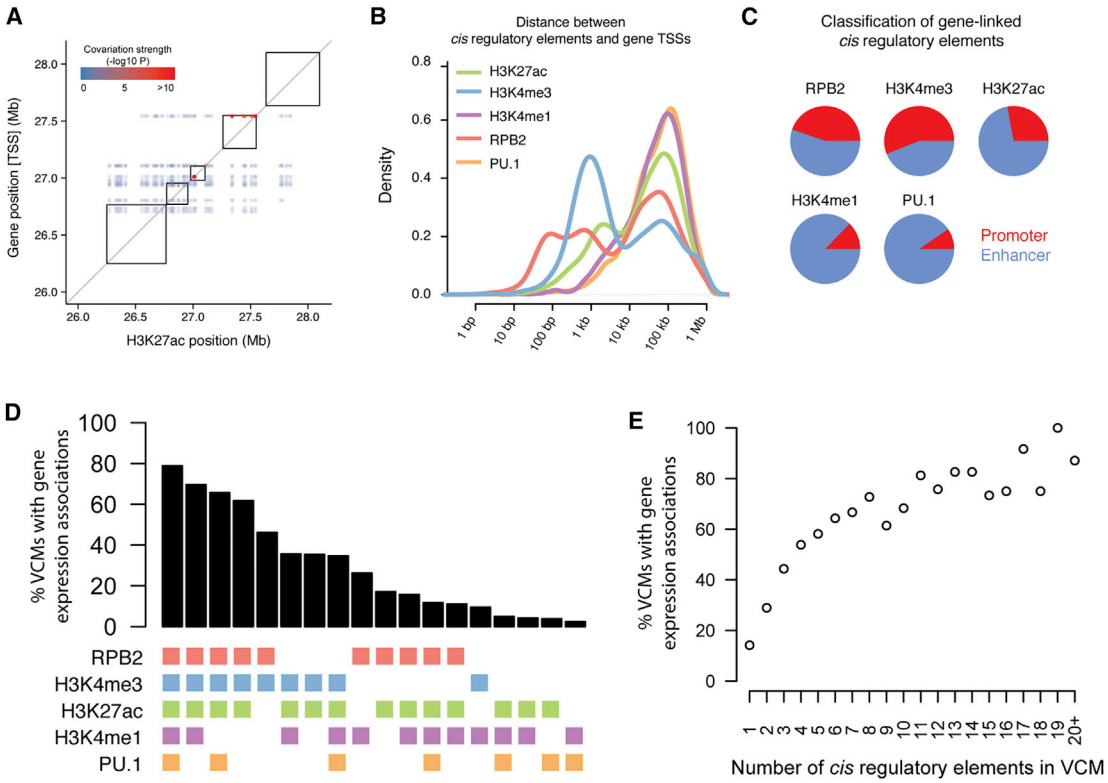


Figure 3. Association between Chromatin State and Gene Expression Variation

(A) Inter-individual co-variation between mRNA levels and H3K27ac enrichment signals at *cis*-regulatory elements on chromosome 21 (26,000,000–28,000,000). The pairwise association strength (Pearson's p value) is color coded and ranges from blue (p = 1) to red (p < 1e-10) (legend, see Figure 1A). Chromosomal contact domains (Rao et al., 2014) are shown with black boxes.

(B) Distance distribution in log-space between the transcription start site (TSS) and *cis*-regulatory elements with expression-linked molecular phenotypes.

(C) Classification of gene-expression-linked *cis*-regulatory elements with molecular marks into putative enhancers and promoters (TSS ± 5 kb).

(D) Percentage of VCMs with gene expression associations (using the first principal component for VCM states) stratified by molecular mark compositions.

(E) Percentage of VCMs associated with gene expression stratified by VCM size (i.e., number of *cis*-regulatory elements that belong to a VCM).

See also Figure S3.

4 and S4). Of note, the number of discovered QTLs significantly increased upon reduction of the *cis*-window size yet at the expense of excluding distal effects (Figures S4A–S4C). Indels and structural variants were significantly enriched among cQTLs (Figure S5A), consistent with previous studies (Kasowski et al., 2010; Schlattl et al., 2011). We further used allele-specific analysis to validate cQTLs on a genome-wide scale (Lappalainen et al., 2013). We observed more significant allelic chromatin biases at cQTLs as compared to control sites (Figure 4C) and higher proportions of allelic chromatin biases at strong cQTLs (Figure 4D), thus supporting our cQTL inference. In addition, we mapped 1,173 vcmQTLs (8.1%) using the first PC as a quantitative trait (comprising 4,187 individual molecular phenotypes) and, surprisingly, none using the second PC despite observing a small enrichment of low p values (Figure S4G). This suggests that the first VCM state captures the primary genetic contributions toward VCM activity. Overall, we found that all molecular phenotypes and, in particular, VCMs are affected by common genetic variants, supporting the hypothesis that a substantial proportion of coordinated chromatin state variation is driven by *cis*-acting genetic variation.

We further assessed the genomic location of cQTLs by measuring their distance relative to TF-targeted and histone-modified regions. We found that the resulting distances exhibit bimodal log-normal distributions, with the first mode centering very close to the mid-point of TF-bound sites (medians between 10 and 40 bp) and relatively close to the mid-point of HM regions (medians between 230 and 300 bp), and the second mode being located distally from its respective target region (medians between 20 and 30 kb) (Figure 4A). In contrast, when we tested the distance relative to the closest TSS (Figure S5B), the log-normal bimodal signal completely disappeared, suggesting that the first mode derives from cQTLs falling within their respectively TF or HM-enriched target regions (Figures S5C and S5D).

Although the proximally mapping cQTLs exhibited significantly stronger effect sizes than cQTLs located outside of their target elements (Figure 4B), they constituted only a minority (25%) of all cQTLs. For example, we found that only 33% of PU.1 QTLs mapped inside of PU.1-bound regions, demonstrating that TF binding is strongly influenced by distal genetic effects. This complexity indicates that, like gene expression, sequence-specific TF-DNA binding can be considered as a

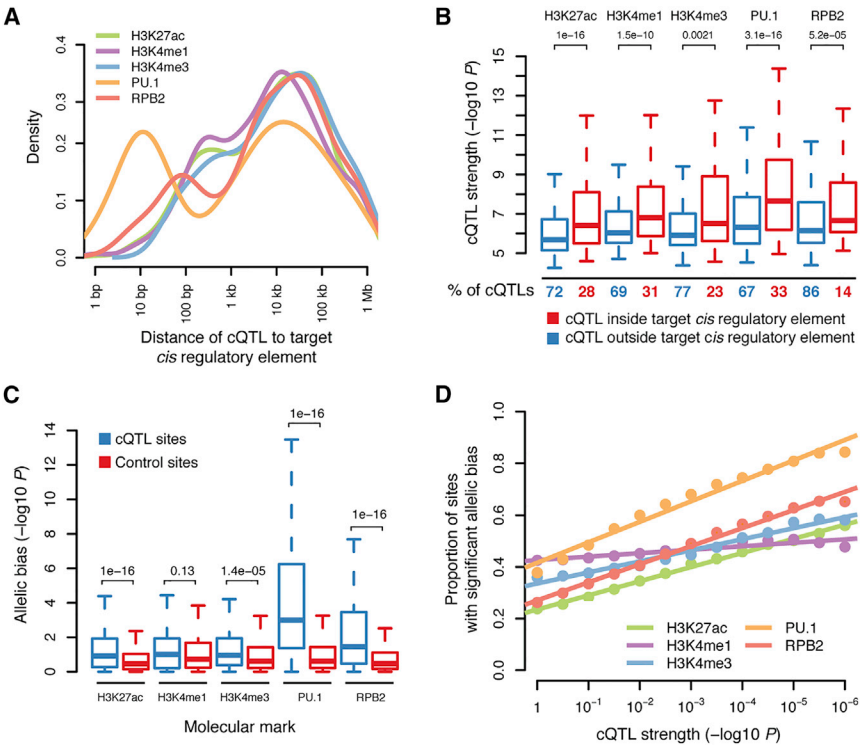


Figure 4. Genetic Control of Chromatin State Variation

(A and B) Quantitative trait locus (QTL) mapping for TF-DNA binding and HMs. (A) Bimodal distance distribution (in log10-space) between cQTLs and their associated *cis*-regulatory elements (FDR 10%). (B) Relationship between cQTL strength and genomic architecture. Boxplots demonstrate genetic association strength (−log10 p value) for QTL variants that map inside (red) and outside (blue) of their target TF-bound/histone-modified regions. Percentages refer to the proportion of cQTLs that fall inside and outside of their target regions.

(C and D) Allele-specific (AS) signals at cQTLs. (C) AS effect strength (−log10 binomial p value) at heterozygous QTL (blue) and non-QTL variants (red). (D) Estimated frequency of AS effects (using π_1 statistics) at heterozygous variants as a function of cQTL strength (−log10 p value). For example, 83% of the heterozygous variants exhibit AS signals in PU.1 binding when considering genetic variants that are associated with PU.1 binding variation at $p < 10^{-6}$.

See also Figures S4 and S5.

complex trait, similar to other molecular and organismal traits. Moreover, we found that distally acting cQTLs exhibited distances that matched the extent of coordination within VCMs, further supporting interactions across regions in the genome. These observations suggest a dual mode of action for cQTLs: strong cQTLs directly perturbing the proximal interactions that form the local chromatin signal and more abundant yet weaker *cis*-acting cQTLs exerting their effects over large distances (up to hundreds of kilobases). The latter process likely involves several intermediate molecular processes that operate within the same VCM.

Given the high degree of quantitative coordination between chromatin state components of the same VCM, we assessed whether distinct molecular phenotypes were affected by the same cQTL. We estimated that half of all cQTLs are shared between two chromatin phenotypes, revealing a strong genetic basis for coordinated chromatin state variation across individuals (Figure 5A). In addition, we found that cQTL-eQTL sharing ranged from a relatively moderate (24% of PU.1 QTLs were also eQTLs) to a very high (73% of H3K4me3 QTLs were also eQTLs) degree (Figure 5A). These results demonstrate that only a small proportion of genetically variable TF-DNA binding events actually lead to measurable changes in gene expression, in line with recent TF knockdown studies carried out in LCLs (Cusanovich et al., 2014). They also suggest that promoter QTLs show very high specificity for genetic gene perturbations, consistent with the enrichment of complex trait-associated variants in cell-type-specific H3K4me3 regions (Trynka et al., 2013).

We further characterized the width and the depth of the QTL signal path by estimating the number of distinct molecular marks and phenotypes that were affected by the same cQTL and eQTL. We observed that the majority of QTLs affect several molecular marks (75%) (Figuress 5B and S6A) and molecular phenotypes of the same and/or different type (80%) (Figures 5C and S6B). Instances of QTLs for which we did not identify cross-talk between distinct molecular marks were of significantly lower effect sizes (Figure S6C). In contrast, 99% of vcmQTLs were associated with multiple molecular marks and phenotypes, suggesting that they capture the deepest and widest range of genetic associations across all studied epigenomic components. Taken together, these results demonstrate that the majority of cQTLs perturb several chromatin state components at the same or across distinct *cis*-regulatory elements.

We next set out to identify which component is more likely to initiate the genetically induced molecular cascade. To do so, we estimated the enrichment of each QTL class being located within particular functional elements. The underlying reasoning was that QTLs that overlap a functional element should initially affect that element first before their effect extends toward non-overlapping elements that belong to the same VCM. We found a clear enrichment signal in TF-bound regions for all types of QTLs. For instance, H3K27ac and H3K4me1 QTLs were seven times more likely to be located within PU.1-bound regions than expected by chance, and vcmQTLs were nine times more enriched within PU.1-bound regions (Figure 5D). We independently validated this observation by testing for enrichment of QTLs in open chromatin regions and experimentally defined TF-bound regions (Figures S6D and S6E). We found that vcmQTLs demonstrated the strongest enrichment at regions that were bound by PU.1, BATF, BCL11A, NFKB, MEF2A, and IRF4 (Figure S6E), consistent with our observations that these TFs are specifically enriched at variable *cis*-regulatory elements (Figure 2E). Moreover, cQTLs that fell within TF-bound regions exhibited stronger

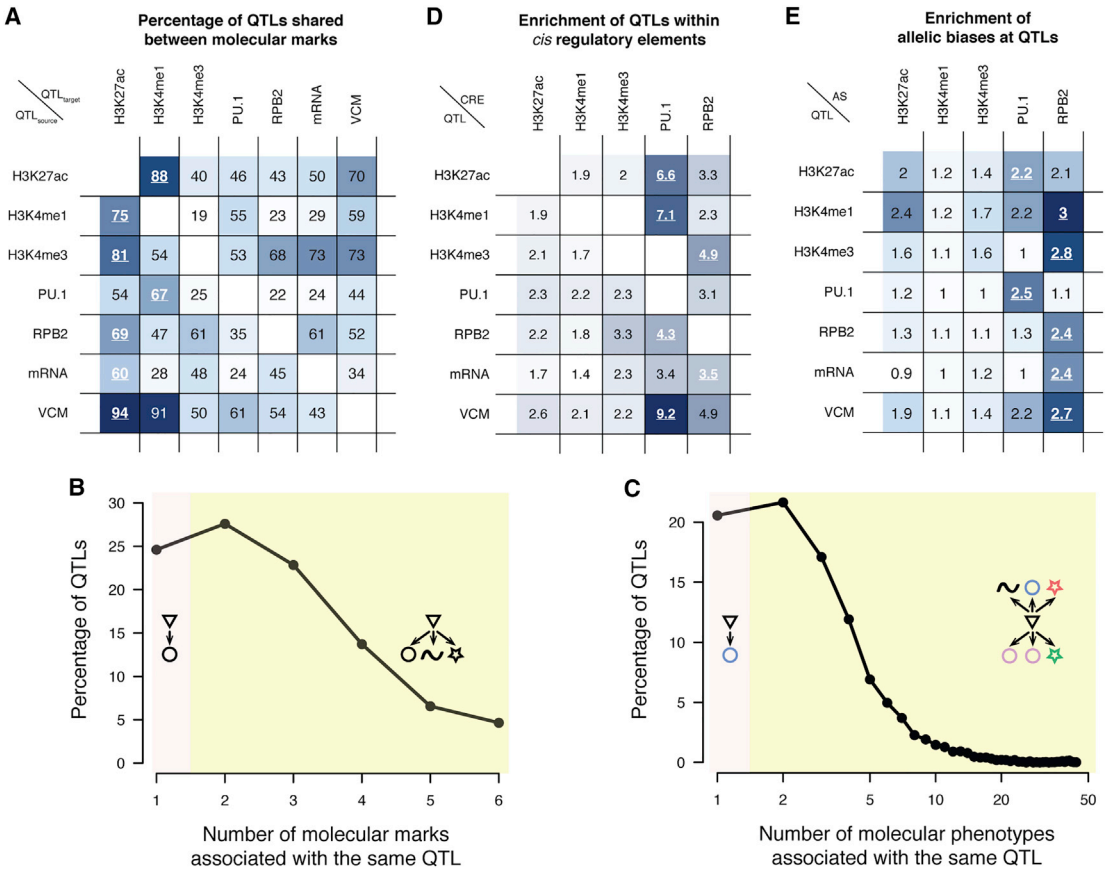


Figure 5. Sharing of Genetic Associations between TF-DNA Binding, HMs, and Gene Expression

(A) Estimation of QTLs that are shared between molecular marks. For example, 81% of H3K4me3 QTLs are also associated with H3K27ac marks.

(B and C) Collateral impact of genetic variation on chromatin architecture and gene expression. (B) Percentage of tf-, hm-, and eQTLs being associated with multiple distinct molecular marks, i.e., DNA binding (PU.1, RPB2), HM levels (H3K4me1, H3K4me3, H3K27ac), and gene expression. For example, 75% of QTLs affect multiple marks. Triangle, genetic variant; other symbols, molecular marks. (C) Percentage of tf-, hm-, and eQTLs being associated with multiple molecular phenotypes (i.e., TF-binding, HM levels, and gene expression). For example, 7.5% of all QTLs affect >10 molecular phenotypes. Triangle, genetic variant; other symbols, molecular phenotypes.

(D) Enrichment of QTLs within active *cis*-regulatory elements. For example, vcmQTL variants map nine times more likely into PU.1-bound regions than expected by chance.

(E) Estimation of allelic effect frequency (using π_1 statistics) at heterozygous QTL variants. For example, AS effects at H3K27ac sites are 2.2-fold more likely at PU.1 QTL variants as compared to all variants.

See also Figures S5 and S6.

effect sizes than those falling outside of such regions (Figure S6F), and we observed stronger enrichment of allelic biases at tfQTLs as compared to hmQTLs for each studied molecular mark (Figure 5E).

We next investigated the impact of TF motif disruption and its downstream effects onto other molecular phenotypes, using Bayesian network modeling. We assessed all molecular associations that involve PU.1 and considered cases separately whereby PU.1 QTL variants disrupted a PU.1 binding site. We observed that PU.1-DNA binding variation was more likely to be causal to variation in H3K27ac and H3K4me1 signals in cases in which the PU.1 motif was disrupted as compared to cases in which the PU.1 QTL mapped elsewhere in the genome (Figure S6G). Thus, these results indicate that sequence-specific TF-DNA interactions are an important driving force behind inter-individual chromatin state variation.

The previous sections demonstrated that genetic perturbation of a few molecular phenotypes can be causal to changes in downstream molecular phenotypes, thus providing a potential explanation as to why most variation in chromatin state is likely independent of proximal sequence changes. VCMs provide the conceptual framework to test the hypothesis of a few molecular phenotypes causing collateral changes to chromatin states across *cis*-regulatory elements. We therefore performed association analysis of vcmQTL variants with every molecular phenotype of the corresponding VCM and observed strong association signals with individual molecular phenotypes (Figures 6A and 6B). Moreover, we observed that the average QTL strength for individual molecular phenotypes scales significantly with the strength of vcmQTLs (ρ = 0.91, p < 2.2e-16) yet one order of magnitude weaker (Figure S6H). The latter observation suggests one or more of the following possibilities: (1) higher-order

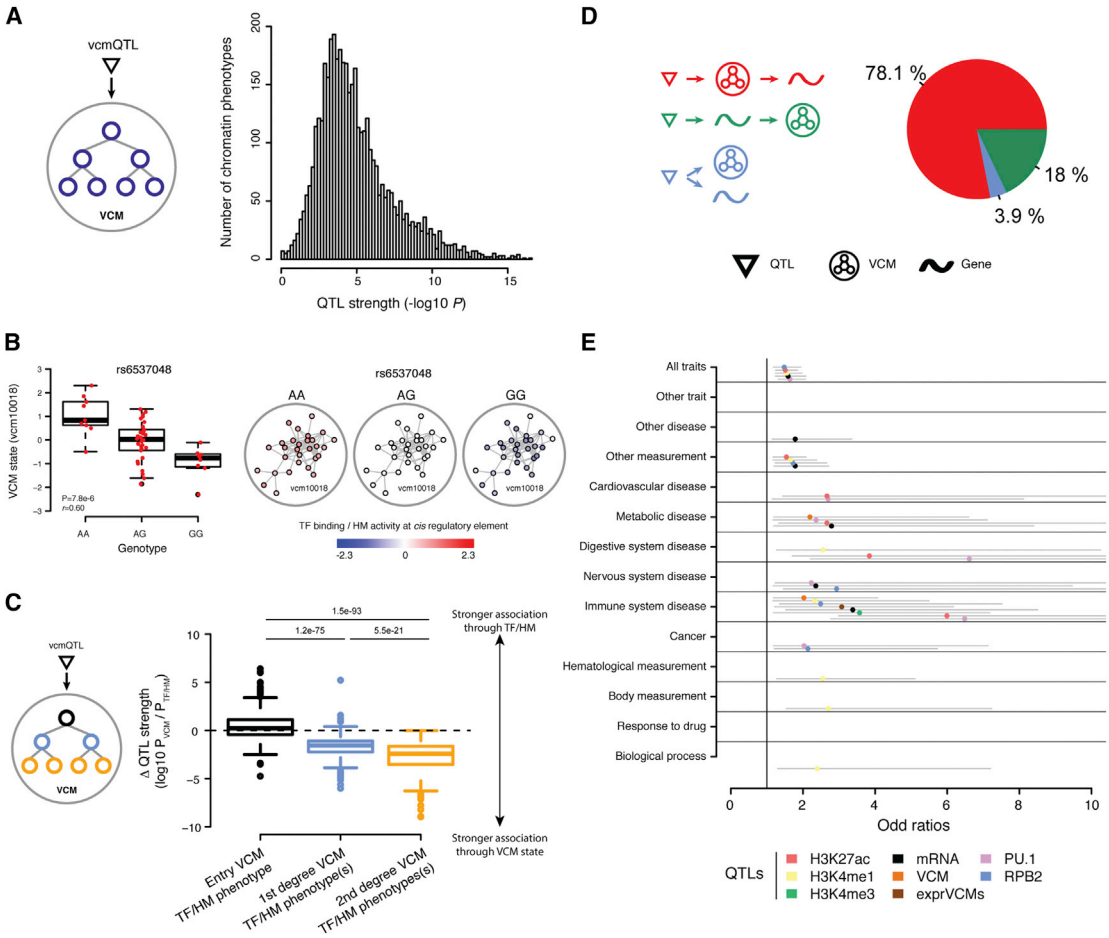


Figure 6. Propagation of Genetic Signals through Molecular Phenotypes

(A) Distribution of −log-transformed association p values for vcmQTL variants and VCM-defining molecular phenotypes.

(B and C) Genetic variation exhibits direct and indirect effects on chromatin architecture. (B) Significant association between the SNP rs6537048 and the state of VCM vcm10018 (chr4:142,224,793–142,570,395) upstream of *IL15*. See Figure S1I for molecular associations in this region based on all marks. Boxplot shows the PCA-derived vcm10018 activity level split by genotype of the SNP rs6537048. Molecular phenotypes within vcm10018 are themselves associated with rs6537048. Molecular association structure is shown together with rs6537048 genotype-averaged TF DNA binding and HM signals. Nodes define individual marks for specific molecular phenotypes (i.e., TF binding and HM) and gray lines depict significant associations between these molecular phenotypes. (C) VCM associations are contrasted against the association strength of the same vcmQTL variant with individual molecular phenotypes (i.e., TF-DNA binding and HM). The molecular association structure within VCMs is used to define three layers of molecular events (entry, first degree, and second degree, see main text). Boxplots show the ratio of genetic association strength between VCMs and the average of individual molecular phenotypes (i.e., log10 $P_{VCM}/P_{TF/HM}$).

(D) Inference of causal relationships between VCM state and gene expression using Bayesian causality modeling. The frequency of the most likely model is shown.

(E) Enrichment of cQTLs and eQTLs in complex disease susceptibility variants by trait class. The gray bars signify 95% CI.

See also Figure S6.

chromatin states are more reflective of genetic perturbations than individual molecular phenotypes; (2) VCMs exhibit a genetically defined structure with few causal effects driving downstream molecular cascades; or (3) VCMs constitute more accurate phenotype estimates, since the correlation structure represented as a PC is devoid of experimental and environmental noise independent of which molecular phenotype is used.

To explore these possibilities, we contrasted the association strength of the same vcmQTL variant with VCM states and individual molecular events (Figure 6C). We further used the molecular association structure that defines VCMs to obtain a hierarchy of molecular interactions: (1) entry phenotypes that exhibit the strongest association with vcmQTLs, (2) direct (first-degree) molecular phenotypes that are defined as being directly associated with the entry phenotype, and (3) indirect (second-degree) molecular interactions that are associated with the entry phenotype via intermediate molecular associations. These analyses revealed that VCM entry phenotypes exhibit a similar association strength with vcmQTL variants as VCMs themselves, further supporting our observation that a single molecular phenotype can act as a seed for collateral changes within the respective VCM (Figure 6C, black boxplot). Interestingly, simulations

demonstrated that PU.1 is most likely to act as an entry phenotype among our probed molecular marks (Figure S6I). Consistent with a hierarchical view, we observed that the remaining molecular phenotypes are, on average, more weakly associated with vcmQTL variants than the overall VCM state and VCM entry phenotypes (Figure 6C, blue and orange boxplots). More specifically, first-degree (direct) molecular phenotypes were more strongly associated with vcmQTL variants than second-degree (indirect) phenotypes.

We subsequently studied genetic variants that affect chromatin modules (vcmQTLs) and gene expression (eQTL) to obtain a global view of the *cis*-regulatory information flow. Bayesian modeling indicates that genetic variants affected gene expression levels through modulation of chromatin activity in 78% of the cases (Figure 6D), thus illustrating that genetic perturbation of chromatin states at *cis*-regulatory elements is, in most cases, causal to changes in gene expression.

Finally, we found that all types of cQTLs are enriched in known complex disease susceptibility variants, especially in immune system disease variants (Figure 6E), providing direct functional genetic evidence that non-coding disease susceptibility variants exert their effects through perturbation of gene regulatory elements.

DISCUSSION

Our analyses uncovered extensive coordination in chromatin variation at and between *cis*-regulatory elements in a human population, revealing the existence of genomic compartments in the form of variable chromatin modules (VCMs). VCMs suggest a higher-order modular organization of gene regulation in the human genome, which is supported by the observation that VCMs are strongly enriched within chromosomal contact domains (Rao et al., 2014). Interestingly, immunity-related genes are specifically associated with VCMs, consistent with the biological nature of LCLs. This finding implies that the resolution of topologically associated domains (TADs) that were so far detected (de Laat and Duboule, 2013; Dixon et al., 2012; Rao et al., 2014) may extend to the level of individual genes (or sets of co-regulated genes), consistent with the observation of micro-topologies at the sub-Mb scale around key developmental genes in mouse embryonic stem cells (Phillips-Cremins et al., 2013). Our data further suggest that population-level chromatin profiling might be an efficient strategy to assess putative chromatin interactions at high spatial resolution, complementing other molecular techniques aimed at mapping chromatin interactions (Sanyal et al., 2012), transcription-coupled chromatin remodeling events (Smolle and Workman, 2013), TF-DNA-binding-induced spreading of histone marks (Hathaway et al., 2012), and enhancer/promoter-gene interactions, respectively.

VCMs also provide a rational framework for explaining why regulatory events can vary independent of proximal sequence changes in molecular terms (Kasowski et al., 2010, 2013; Kilpinen et al., 2013; McVicker et al., 2013; Reddy et al., 2012; Villar et al., 2014). Chromatin activity at *cis*-regulatory elements can be influenced by distally acting genetic variants of variable effect size, as we strongly suggest in this study for all analyzed molecular phenotypes. In addition, we found that the activity level of each VCM can be captured by a single quantitative phenotype, which suggests that molecular processes within each VCM (i.e., histone-mark deposition and TF-DNA binding) are subject to highly penetrant causal events. Our study provides strong support for the hypothesis that these events correspond in large part to genetic perturbations of TF-DNA interactions. This is based on the fact that vcmQTLs: (1) are strongly enriched within TF-occupied regions, (2) simultaneously perturb several layers of chromatin structure, and (3) are in the majority of cases causal to the observed changes in gene expression. From this, a model emerges in which the perturbation of a single or a few TF-DNA interactions can act as a seed for coordinated, collateral regulatory changes within a respective VCM. We hypothesize that these changes are in large part mediated by long-range TF-TF cooperativity events, given our observation that specific pairs of lineage-determining, signal-dependent, and architectural factors (Heinz et al., 2010; Ong and Corces, 2014; Zhou et al., 2015) are significantly enriched at VCM elements.

Interestingly, whereas ''promoter VCMs'' correlated frequently with gene expression, we found that only a few ''enhancer VCMs'' were linked to nearby genes, and only one-quarter of PU.1 or H3K4me1 QTLs were shared with eQTLs. This finding may imply either: (1) that abundant enhancer variation is of such small effect on target gene expression as to be undetectable given the sample size of this study or (2) that the affected enhancers are primed to conditionally regulate gene expression (for example, in response to specific stimuli) (Calo and Wysocka, 2013; Shlyueva et al., 2014; Spitz and Furlong, 2012). Alternatively, these sequences may be subject to spurious regulatory activity, which would explain the findings that: (1) only a minority of genetically variable TF-DNA binding events result in differential gene expression (this work), (2) a large portion of TF-DNA binding events have no functional consequence (Cusanovich et al., 2014; Farnham, 2009), and (3) TF binding sites tend to experience rapid turnover (Dermitzakis and Clark, 2002; Villar et al., 2015). Another complementary interpretation involves the model of dose-dependent gene activation in which several TF binding sites in multiple elements cumulatively contribute to gene expression (Spivakov, 2014). Under this model, loss of TF-DNA binding at one binding site would have little to no discernible functional consequence as long as the other implicated TF binding sites remain intact. This would, in turn, be consistent with our observation that VCMs involving multiple *cis*-regulatory elements were far more likely to correlate with gene expression variation than VCMs involving only one element.

Our present work on the discovery of molecular associations and cQTLs for key chromatin organization components in a human population sample provides unique insights and a novel framework for studying the molecular mechanisms underlying variable transcriptional programs between individuals.

EXPERIMENTAL PROCEDURES

Study Samples

ChIP-seq and RNA-seq data were produced from lymphoblastoid cell lines (LCLs) of 54 samples (Abecasis et al., 2010). All individuals were of European origin (Utah residents with ancestry from northern and western Europe and abbreviated as CEU). After excluding samples due to suspected swaps, contamination (see Supplemental Experimental Procedures, 3.4), or

incomplete data availability (sample failed for a subset of assays), our final data set consisted of 47 individuals for all ChIP assays and 46 individuals for gene expression measurements (Table S1 for basic sample information).

ChIP-Seq and mRNA-Seq Experiments

All sequencing assays (ChIP and mRNA) were produced from a single growth of LCLs, and cell culture and cell fixation were performed as previously described (Kilpinen et al., 2013). ChIPs for H3K27ac, H3K4me1, H3k4me3, PU.1, and RNA polymerase II (RPB2) were performed as described in the Supplemental Experimental Procedures, 1.1–1.3. RNA extraction was done following the procedure described in the Supplemental Experimental Procedures, 2.1. Library preparation and sequencing done for ChIP and mRNA are described in detail in Supplemental Experimental Procedures, 1.4 and 2.2, respectively. Short-read alignment for ChIP-seq and RNA-seq was performed using BWA 0.5.9 (Li and Durbin, 2009) against the hg19 build of the human reference genome supplemented with the Epstein-Barr virus (EBV) sequence. All sequencing data management was done using Samtools (Li et al., 2009) (Supplemental Experimental Procedures, 1.5 and 2.3). A summary of mapping statistics is provided in Table S1B.

From ChIP-Seq Experiments to Molecular Phenotypes

ChIP-seq peak calling was not directly performed in the current set of samples to avoid the issue of fuzzy peak boundaries. Instead, we used an independently derived peak set for each assay that is based on six 1000 Genomes Project Pilot individuals (Kilpinen et al., 2013). Quantifications for all peak-sample pairs were obtained by counting overlapping reads using HOMER (Heinz et al., 2010), which resulted in a quantification matrix of size #samples × #peaks per assay (Supplemental Experimental Procedures, 1.6). Peak quantifications were scaled to adjust for differences in total library size and were corrected for batch effects using PEER (Stegle et al., 2010). We empirically determined the optimal number of K PEER factors to be removed by finding the K leading to the highest number of discovered QTLs (Supplemental Experimental Procedures, 1.7).

From mRNA-Seq Experiments to Molecular Phenotypes

mRNA-seq data were quantified per sample based on GENCODE v15 (08/2012) gene annotations (Harrow et al., 2012), resulting in a quantification matrix of size #samples × #genes. All genes with five samples (>10%) or more without any overlapping reads were removed, and the remaining quantifications were scaled (10 M reads) and corrected for batch effects (PEER K = 15) (Supplemental Experimental Procedures, 2.4 and 2.5).

Genotype Information

Genotypes for the 47 samples were obtained from two sources: (1) 34 with genome-wide sequence data from 1000 Genomes Phase 1 v.3 and (2) 13 other CEU samples with Illumina Omni2.5 genotype data. Both were merged by imputing untyped sequence variants into Illumina Omni2.5 data using IMPUTE2 (Howie et al., 2009). Subsequently, all variants with minor allele frequency below 5% were removed. See Lappalainen et al. (2013) for additional details on genotype processing.

Analytical Methods for Molecular Phenotype-Phenotype Associations

Mapping Molecular Associations

To map associations between pairs of peaks, we proceeded as follows for each of the 15 possible unordered pairs of distinct molecular phenotypes: we measured the inter-individual Pearson correlation and its significance (p value) between quantifications of every possible pair of peaks within 1 Mb distance of each other. Then, we corrected for multiple testing by controlling for a 0.1% false discovery rate using the R/*qvalue* package (Alan Dabney, John D. Storey and with assistance from Gregory R. Warnes. qvalue: Q-value estimation for false discovery rate control. R package version 1.36.0.). Percentages (i.e., π_1 estimates) of truly associated pairs were also estimated as a byproduct (Supplemental Experimental Procedures, 3.1).

Building VCMs

We used graph theory to build VCMs and assumed that peaks are nodes and significant peak associations edges. Any two peaks belong to the same VCM as soon as there is a path (i.e., a sequence of edges) that links them together; otherwise, they belong to two distinct VCMs. Based on this, we implemented an iterative algorithm that assigns peaks to VCMs. Then, VCM state activity levels were obtained using principal component analysis (PCA) on quantifications of all peaks that belong to a VCM (Supplemental Experimental Procedures, 3.2).

Functional Annotation of VCMs

We used the online service *GREAT* v2.0.2 to predict over-represented pathways and biological processes for VCM domains. Functional annotation of VCM-associated genes was performed using the online service *ConsensusPathDB-human* using the over-representation analysis module and gene ontology categories (BP level 2) (Supplemental Experimental Procedures, 3.9).

Enrichment in Contact Domains

We used high-resolution chromosomal contact domains for LCLs from Rao et al. (2014) to estimate how more likely associated peak pairs occur within the same contact domain as compared to non-associated ones. To do so, we used logistic regression with within/between contact domains as the binary response, the association status (significant or not) as explanatory variable, and the peak-to-peak distance as a covariate (Supplemental Experimental Procedures, 3.11).

TFs Co-Occurrence at VCM Elements

We used the Fisher's exact test to estimate enrichments of ENCODE TF-TF pairs at non-overlapping VCM elements (Supplemental Experimental Procedures, 3.13).

Analytical Methods for Quantitative Trait Loci

Mapping QTLs

We mapped *cis*-acting quantitative trait loci (QTLs) by performing linear regressions between peak or gene quantifications and genotypes at all variant sites within 250 kb (*cis*-window around the gene TSS or the peak center). Then, we stored the best association for each peak/gene as a putative QTL and corrected: (1) for multiple variants and (2) multiple peaks/genes being tested genome-wide. We used permutations and false discovery rate to correct for (1) and (2), respectively. In addition, we repeated this analysis multiple times with various *cis*-window sizes in order to determine the size providing the best trade-off between discovery power and distal effect capture (Figures S4A–S4C and Supplemental Experimental Procedures, 3.3). This analysis has been performed using the software package FastQTL (http://fastqtl.sourceforge.net/).

Estimating Proportion of Shared QTLs

To see if a QTL for assay A1 is replicated in assay A2, we first found a A2 peak that matches the A1 peak by minimizing the distance between both, and then we looked at the nominal p value of association between the QTL and the matched A2 peak. By repeating this for all A1 QTLs, we can then estimate the proportion that is shared with A2 using the π_1 statistic (Supplemental Experimental Procedures, 3.5).

Detecting Multiple Effects of QTLs

To map out the peaks affected by a QTL, we measured association between the QTL and all features across all assays located within 250 kb and then divided the resulting p values by the number of tested features (Bonferroni correction) and finally reported as hits associations with a p value below the 0.05 threshold (Supplemental Experimental Procedures, 3.6).

Enrichment of QTLs within Functional Annotations

To measure how much more likely than by chance a set of QTLs is located within a particular annotation, we developed an approach that corrects for the fact that QTLs and annotations are not uniformly distributed along the genome, with the goal of getting more robust enrichment estimates. This method basically aims to find a null set of QTLs with some properties (e.g., distance to associated peak/gene) that match the original set (Supplemental Experimental Procedures, 3.7).

Enrichment of QTLs with GWAS Hits

To measure how the QTL sets are enriched for GWAS hits, we used the NHGRI GWAS Catalog (December 8, 2014), generated 1,000 null sets of QTLs with matching properties (distance to associated peak/gene and minor allele frequency), and tested how often these null QTL sets overlap GWAS hits as compared to the original QTL set. Note that two variants are assumed to overlap as soon as they are in high LD (Supplemental Experimental Procedures, 3.10).

QTL Causality Modeling

When a QTL is associated with two peaks (or genes), we inferred the most likely signal transmission path (i.e., the causal chain of events) through the two affected molecular phenotypes using Bayesian network modeling: we enumerate the three possible models (QTL = > A1 = > A2, QTL = > A2 = > A1 and QTL = > A1/QTL = > A2), estimate their respective likelihood, and assign the most likely model to each triplet (Supplemental Experimental Procedures, 3.8).

Analytical Methods for Allele-Specific Effects

Mapping ASE

This was only performed on samples with sequence data (n = 34/47, Experimental Procedures, "Genotype Information") at heterozygous SNPs. Deviation from equilibrium (i.e., 50%–50%) was characterized using binomial tests, accounting for multiple major sources of technical bias, such as reference allele mapping bias, clonal reads, and non-unique mappability of reads, as described previously (Kilpinen et al., 2013; Lappalainen et al., 2013; Waszak et al., 2014) (Supplemental Experimental Procedures, 3.4). Allele-specific effects (ASE) analysis was also used as a quality control step to identify putative sample swaps or contaminations.

Haplotypic ASE Coordination

We looked at ASE measured at phased heterozygous SNPs falling within VCM peaks and assessed whether the signal was consistent with the haplotype phase. In practice, we use logistic regression with concordance in allelic direction as response variable, association status (VCM/null) as explanatory variable, and distance between peaks as covariates (Supplemental Experimental Procedures, 3.12).

ACCESSION NUMBERS

All BAM files for this study have been submitted to the ArrayExpress Archive (http://www.ebi.ac.uk/arrayexpress/). The accession numbers are: E-MTAB-3656 (mRNA-seq data) and E-MTAB-3657 (ChIP-seq data).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.08.001.

AUTHOR CONTRIBUTIONS

E.T.D., B.D., A.R., and N.H. designed and supervised the study and contributed to data interpretation. S.T. and D.H. cultured and processed the lymphoblastoid cell lines. S.K.R., R.M.W., and A.O. designed the ChIP experiments. S.K.R., R.M.W., A.O., M.W., and G.U. executed the ChIP experiments. L.R.-P., A.P., and D.B. prepared ChIP- and RNA-seq libraries and executed the ChIP- and RNA-seq experiments. S.M.W. and O.D. designed and executed the primary data analysis. A.R.G. with O.D. designed and executed the causality inference analysis. H.K., N.I.P., A.Y., and I.P. contributed to data analysis. S.M.W., O.D., B.D., and E.T.D. performed the primary manuscript writing with contributions from N.H. and A.R.G.

ACKNOWLEDGMENTS

Supported by Swiss National Science Foundation grants CRSI33_130326 (E.T.D., B.D., A.R., N.H.), 31003A_132958 (N.H.), 31003A_130342 and 149984 (E.T.D.), 31003A_129835 (A.R.), and 31003A_138323 (B.D.); the European Research Council (E.T.D.); the Louis-Jeantet Foundation (E.T.D.); SystemsX grant 3826 (E.T.D.); NIH MH101814 (E.T.D.); European Molecular Biology Organization fellowship ALTF 2010-337 (H.K.); a fellowship from the doctoral school of the Faculty of Biology and Medicine, University of Lausanne (R.M.W.); the NCCR Frontiers in Genetics Program (E.T.D., B.D.); the Japanese-Swiss Science and Technology Cooperation Program (Japan Science and Technology Agency/ETH Zürich) (B.D.); and École Polytechnique Fédérale de Lausanne (B.D.). The computations were performed at the Vital-IT (http://www.vital-it.ch) Center for High-Performance Computing of the Swiss Institute of Bioinformatics.

Received: October 1, 2014
Revised: March 17, 2015
Accepted: July 29, 2015
Published: August 20, 2015

REFERENCES

Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., McVean, G.A., and McVean, G.A.; 1000 Genomes Project Consortium (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061–1073.

Albert, F.W., and Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. Nat. Rev. Genet. *16*, 197–212.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol. Cell *49*, 825–837.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Cusanovich, D.A., Pavlovic, B., Pritchard, J.K., and Gilad, Y. (2014). The functional consequences of variation in transcription factor binding. PLoS Genet. *10*, e1004226.

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

Dermitzakis, E.T., and Clark, A.G. (2002). Evolution of transcription factor binding sites in Mammalian gene regulatory regions: conservation and turnover. Mol. Biol. Evol. *19*, 1114–1121.

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.

Fairfax, B.P., Makino, S., Radhakrishnan, J., Plant, K., Leslie, S., Dilthey, A., Ellis, P., Langford, C., Vannberg, F.O., and Knight, J.C. (2012). Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat. Genet. *44*, 502–510.

Farnham, P.J. (2009). Insights from genomic profiling of transcription factors. Nat. Rev. Genet. *10*, 605–616.

Grundberg, E., Small, K.S., Hedman, Å.K., Nica, A.C., Buil, A., Keildson, S., Bell, J.T., Yang, T.-P., Meduri, E., Barrett, A., et al.; Multiple Tissue Human Expression Resource (MuTHER) Consortium (2012). Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat. Genet. *44*, 1084–1089.

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.

Hathaway, N.A., Bell, O., Hodges, C., Miller, E.L., Neel, D.S., and Crabtree, G.R. (2012). Dynamics and memory of heterochromatin in living cells. Cell *149*, 1447–1460.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Heinz, S., Romanoski, C.E., Benner, C., Allison, K.A., Kaikkonen, M.U., Orozco, L.D., and Glass, C.K. (2013). Effect of natural genetic variation on enhancer selection and function. Nature *503*, 487–492.

Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. *5*, e1000529.

Karczewski, K.J., Tatonetti, N.P., Landt, S.G., Yang, X., Slifer, T., Altman, R.B., and Snyder, M. (2011). Cooperative transcription factor associations discovered using regulatory variation. Proc. Natl. Acad. Sci. USA *108*, 13353–13358.

Kasowski, M., Grubert, F., Heffelfinger, C., Hariharan, M., Asabere, A., Waszak, S.M., Habegger, L., Rozowsky, J., Shi, M., Urban, A.E., et al. (2010). Variation in transcription factor binding among humans. Science *328*, 232–235.

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.

Kilpinen, H., Waszak, S.M., Gschwind, A.R., Raghav, S.K., Witwicki, R.M., Orioli, 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.

Lappalainen, T., Sammeth, M., Friedländer, M.R., 't Hoen, P.A., Monlong, J., Rivas, M.A., Gonzàlez-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, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Manolio, T.A. (2010). Genomewide association studies and assessment of the risk of disease. N. Engl. J. Med. *363*, 166–176.

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.

Nica, A.C., Montgomery, S.B., Dimas, A.S., Stranger, B.E., Beazley, C., Barroso, I., and Dermitzakis, E.T. (2010). Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. PLoS Genet. *6*, e1000895.

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.

Ong, C.-T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. Nat. Rev. Genet. *15*, 234–246.

Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.-T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell *153*, 1281–1295.

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.

Reddy, T.E., Gertz, J., Pauli, F., Kucera, K.S., Varley, K.E., Newberry, K.M., Marinov, G.K., Mortazavi, A., Williams, B.A., Song, L., et al. (2012). Effects of sequence variation on differential allelic transcription factor occupancy and gene expression. Genome Res. *22*, 860–869.

Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature *489*, 109–113.

Schlattl, A., Anders, S., Waszak, S.M., Huber, W., and Korbel, J.O. (2011). Relating CNVs to transcriptome data at fine resolution: assessment of the effect of variant size, type, and overlap with functional regions. Genome Res. *21*, 2004–2013.

Shlyueva, D., Stelzer, C., Gerlach, D., Yáñez-Cuna, J.O., Rath, M., Boryń, L.M., Arnold, C.D., and Stark, A. (2014). Hormone-responsive enhancer-activity maps reveal predictive motifs, indirect repression, and targeting of closed chromatin. Mol. Cell *54*, 180–192.

Smolle, M., and Workman, J.L. (2013). Transcription-associated histone modifications and cryptic transcription. Biochim. Biophys. Acta *1829*, 84–97.

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626.

Spivakov, M. (2014). Spurious transcription factor binding: non-functional or genetically redundant? BioEssays *36*, 798–806.

Stefflova, K., Thybert, D., Wilson, M.D., Streeter, I., Aleksic, J., Karagianni, P., Brazma, A., Adams, D.J., Talianidis, I., Marioni, J.C., et al. (2013). Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. Cell *154*, 530–540.

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.

Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. USA *100*, 9440–9445.

Trynka, G., Sandor, C., Han, B., Xu, H., Stranger, B.E., Liu, X.S., and Raychaudhuri, S. (2013). Chromatin marks identify critical cell types for fine mapping complex trait variants. Nat. Genet. *45*, 124–130.

Villar, D., Flicek, P., and Odom, D.T. (2014). Evolution of transcription factor binding in metazoans - mechanisms and functional implications. Nat. Rev. Genet. *15*, 221–233.

Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M., Pignatelli, M., Park, T.J., Deaville, R., Erichsen, J.T., Jasinska, A.J., et al. (2015). Enhancer evolution across 20 mammalian species. Cell *160*, 554–566.

Waszak, S.M., Kilpinen, H., Gschwind, A.R., Orioli, A., Raghav, S.K., Witwicki, R.M., Migliavacca, E., Yurovsky, A., Lappalainen, T., Hernandez, N., et al. (2014). Identification and removal of low-complexity sites in allele-specific analysis of ChIP-seq data. Bioinformatics *30*, 165–171.

Wray, G.A. (2007). The evolutionary significance of cis-regulatory mutations. Nat. Rev. Genet. *8*, 206–216.

Zheng, W., Zhao, H., Mancera, E., Steinmetz, L.M., and Snyder, M. (2010). Genetic analysis of variation in transcription factor binding in yeast. Nature *464*, 1187–1191.

Zhou, H., Schmidt, S.C., Jiang, S., Willox, B., Bernhardt, K., Liang, J., Johannsen, E.C., Kharchenko, P., Gewurz, B.E., Kieff, E., and Zhao, B. (2015). Epstein-Barr virus oncoprotein super-enhancers control B cell growth. Cell Host Microbe *17*, 205–216.