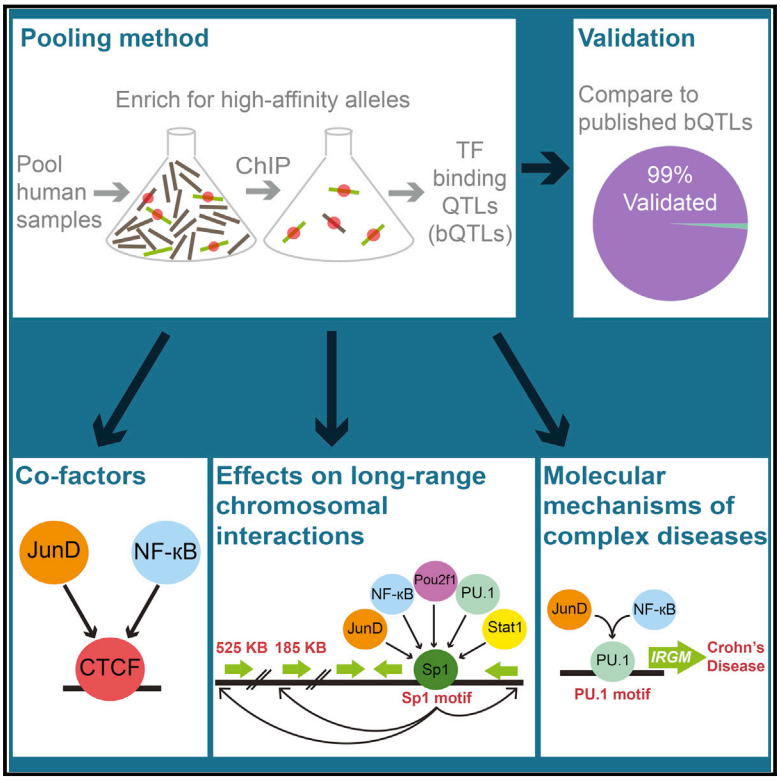
Resource

Cell

Pooled ChIP-Seq Links Variation in Transcription Factor Binding to Complex Disease Risk

Graphical Abstract



Authors

Ashley K. Tehranchi, Marsha Myrthil, Trevor Martin, Brian L. Hie, David Golan, Hunter B. Fraser

Correspondence

hbfraser@stanford.edu

In Brief

Examination of thousands of human genetic variants that affect transcription factor binding demonstrates a role for natural gene variation in chromosomal architecture and illustrates the efficiency and economy of using pooled samples for these analyses.

Highlights

- A pooling-based approach maps QTLs for molecular-level traits with reduced cost
- Thousands of *cis*-acting QTLs affect transcription factor binding in humans
- CTCF anchors binding of multiple transcription factors
- Binding QTLs link genetic variation to 3D genome architecture and complex traits

Tehranchi et al., 2016, Cell *165*, 730–741
April 21, 2016
http://dx.doi.org/10.1016/j.cell.2016.03.041

CellPress

Cell

Resource

Pooled ChIP-Seq Links Variation in Transcription Factor Binding to Complex Disease Risk

Ashley K. Tehranchi,[1] Marsha Myrthil,[2] Trevor Martin,[1] Brian L. Hie,[3] David Golan,[4,5] and Hunter B. Fraser[1,*]
[1]Department of Biology, Stanford University, Stanford, CA 94305, USA
[2]Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA
[3]Department of Computer Science
[4]Department of Genetics
[5]Department of Statistics
Stanford University, Stanford, CA 94305, USA
*Correspondence: hbfraser@stanford.edu
http://dx.doi.org/10.1016/j.cell.2016.03.041

SUMMARY

Cis-regulatory elements such as transcription factor (TF) binding sites can be identified genome-wide, but it remains far more challenging to pinpoint genetic variants affecting TF binding. Here, we introduce a pooling-based approach to mapping quantitative trait loci (QTLs) for molecular-level traits. Applying this to five TFs and a histone modification, we mapped thousands of *cis*-acting QTLs, with over 25-fold lower cost compared to standard QTL mapping. We found that single genetic variants frequently affect binding of multiple TFs, and CTCF can recruit all five TFs to its binding sites. These QTLs often affect local chromatin and transcription but can also influence long-range chromosomal contacts, demonstrating a role for natural genetic variation in chromosomal architecture. Thousands of these QTLs have been implicated in genome-wide association studies, providing candidate molecular mechanisms for many disease risk loci and suggesting that TF binding variation may underlie a large fraction of human phenotypic variation.

INTRODUCTION

In the past decade, genome-wide association studies (GWAS) have transformed our view of common diseases: thousands of loci have been found to associate with common disease risk, compared to just a handful before the GWAS era. However, despite their success, GWAS have suffered from several major limitations (Edwards et al., 2013). First, they generally do not reveal the causal variants underlying the associations, due to linkage disequilibrium (LD) between variants. Second, they give no information about the molecular mechanisms of how these variants affect disease risk; this generally requires laborious follow-up studies of individual loci. As a result, GWAS have made little contribution to our understanding of disease etiologies.

Perhaps the clearest trend to emerge from most GWAS is that the vast majority of genetic variants affecting complex traits resides outside of protein-coding regions (Hindorff et al., 2009; Schaub et al., 2012; Edwards et al., 2013); recent evolutionary adaptations follow a similar pattern (Fraser, 2013; Enard et al., 2014). These trait-associated noncoding variants are highly enriched in transcriptional enhancers (Gusev et al., 2014; Farh et al., 2015; Kundaje et al., 2015), but whether they affect enhancer function—and if so, how—remains largely unknown. One approach to annotate candidate functions for GWAS hits is to intersect them with quantitative trait loci (QTLs) mapped for molecular-level traits such as gene expression (eQTLs). This can be informative, but only in a small fraction of cases (e.g., 5.9% of GWAS hits were in LD with an eQTL from the multi-tissue GTEx analysis) (GTEx Consortium, 2015), and it still does not indicate how these variants affect gene expression.

A complementary approach for annotating noncoding variants is to determine which of them are associated with transcription factor (TF) binding. This may be quite powerful, since TF binding (1) is mostly determined by highly local sequence (White et al., 2013; Levo and Segal, 2014), reducing the search space for QTLs; (2) is causally upstream of transcription and thus may be more strongly linked to genetic variants (cf., Kaplow et al., 2015); and (3) involves many more traits (TF binding sites), and thus potential QTLs, than the number of genes.

Chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) in many genotyped human samples could allow the genetic mapping of variants affecting TF binding, but in practice, this is quite challenging due to both experimental variation between samples (e.g., batch effects) (Bonhoure et al., 2014) and the prohibitive cost of such an experiment. As a result, previous studies comparing TF binding across individuals—that have revealed extensive inter-individual variation—have been limited to relatively small sample sizes (Kasowski et al., 2010, 2013; Kilpinen et al., 2013; Ding et al., 2014; Waszak et al., 2015). In this work, we developed a more efficient and cost-effective approach to mapping QTLs for molecular-level traits, which allowed us to map thousands of QTLs affecting TF binding or histone modification. Their extensive overlap with variants previously implicated by GWAS provides candidate molecular

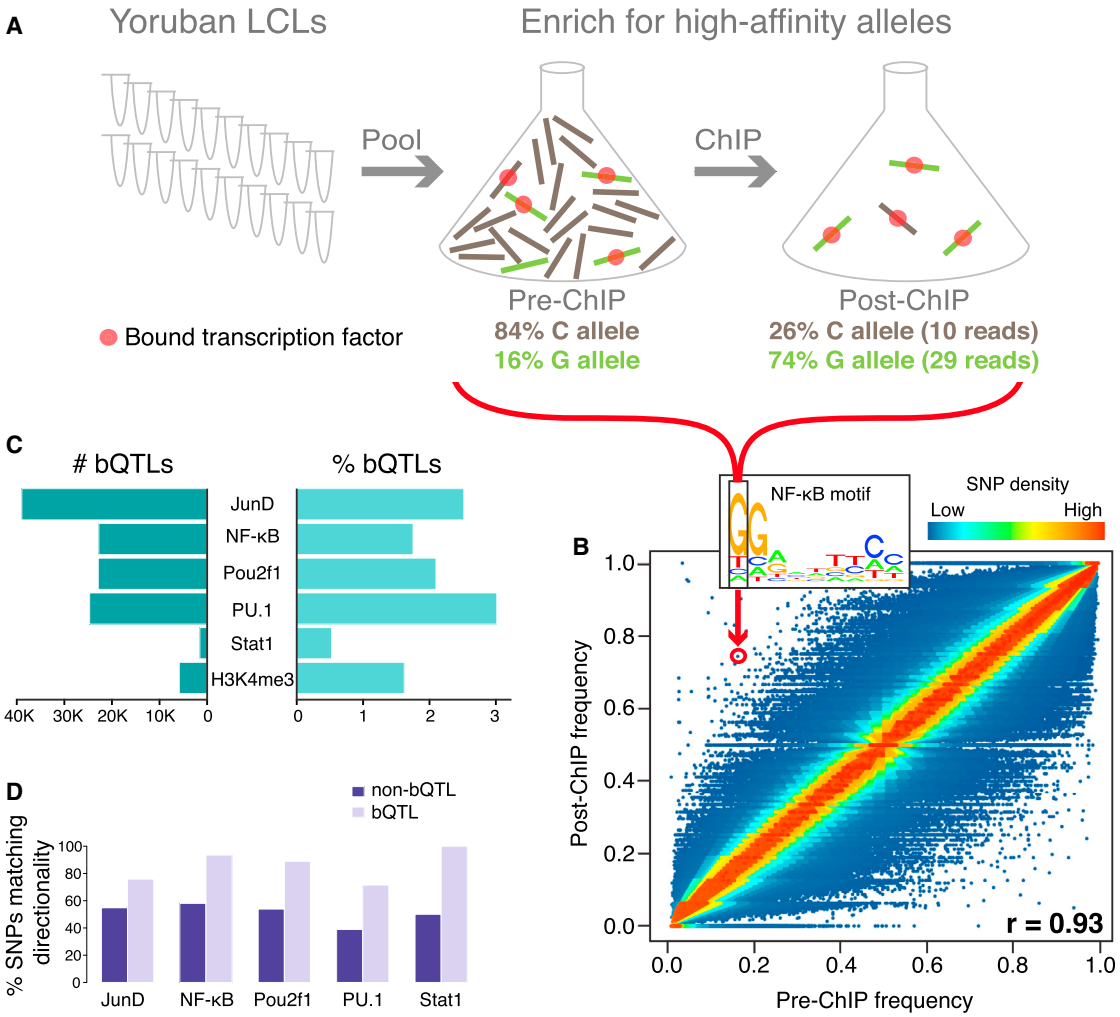


Figure 1. Outline and Results of Pooled ChIP-Seq

(A) Performing ChIP in a pool of individuals selects only the DNA molecules bound by a TF (or with a particular histone modification), thus enriching for high-affinity alleles. In this example, the G allele has a low pre-ChIP frequency but a high post-ChIP frequency, due to its higher affinity. The allele frequencies and read counts are for the SNP highlighted in (B).

(B) For NF-κB, plotting pre-ChIP versus post-ChIP allele frequencies shows that the vast majority of allele frequencies change very little, as expected. One SNP (rs10263017) significantly off the diagonal is highlighted. This SNP occurs in the boxed position of an NF-κB binding site motif, where the G allele has higher predicted affinity, consistent with the allele frequency shift.

(C) The number of independent bQTLs (after removing those in LD) (left) and the percent of all bound SNPs called as bQTLs (right).

(D) Directionality agreement between predicted effects of SNPs in TF binding motifs and their observed associations. SNPs that are bound but have no allele frequency shift ($p > 0.5$) show ~50% agreement, as expected by chance; bQTL ($p < 0.005$) agreement is significantly higher. H3K4me3 has no motif, so is not included. All comparisons are significant at Fisher's exact test $p < 0.005$ except for Stat1, due to its small number of bQTLs.

See also Figure S1 and Table S1.

mechanisms underlying disease etiologies, while more broadly suggesting a major role for TF binding in complex disease risk.

RESULTS

Pooled ChIP-Seq for High-Throughput QTL Mapping

We devised a simple modification to existing methods for mapping QTLs controlling molecular-level traits that reduces both experimental variation and cost. The central idea is that samples from any number of individuals can be combined into a single pool, in which ChIP-seq is performed only once. For genetic variants that affect binding of a TF in *cis*, the high-affinity allele will be present at a higher frequency after the ChIP, as compared to before (Figure 1A). In contrast, variants with no effect on binding should have no significant shift in allele frequencies. Post-ChIP allele frequencies are estimated based on the observed counts of each allele in the ChIP-seq reads. One potential complication is that samples with greater abundance or activity of a TF will have more bound sites and thus contribute more to the post-ChIP pool; therefore, we estimated pre-ChIP allele

frequencies with a regression-based approach that accounts for any differences in TF activities between samples (see the Supplemental Experimental Procedures). Our approach is limited to variation in local sequence (covered by the ChIP-seq reads), which is the primary determinant of TF binding (White et al., 2013; Levo and Segal, 2014).

We performed ChIP-seq for five TFs critical for immune cell development and function (NF-κB, PU.1/Spi1, Stat1, JunD, and Pou2f1/Oct1) in a pool of 60 Yoruban lymphoblastoid cell lines (LCLs) and obtained ~300 million reads per TF (except for Stat1 with ~126 million reads) with excellent agreement between biological replicates (Figure S1A). We also produced ~804 million reads for a histone modification associated with active transcription (H3K4me3) in a pool of 71 Yoruban LCLs. In all experiments, we found that the vast majority of SNPs had similar pre-ChIP and post-ChIP allele frequencies (Figures 1B and S1B), as expected. However, many SNPs had significantly shifted frequencies, assessed by a modified binomial test that accounts for uncertainty in our allele frequency estimates (see the Supplemental Experimental Procedures). We refer to these SNPs as binding QTLs (bQTLs). We found between 1,533–38,944 independent bQTLs per TF at a nominal $p < 0.005$, which represented 0.5%–3% of all bound SNPs (Figs 1C and S1C; Table S1; false discovery rates discussed below).

For example, the G allele of rs10263017 was strongly overrepresented in the NF-κB post-ChIP sample (Figure 1B; $p = 1.8 \times 10^{-15}$), consistent with its being located in an NF-κB binding motif where the G allele has a higher predicted affinity. Similarly, among all of our bQTLs predicted to affect the corresponding TF's motif (see the Supplemental Experimental Procedures), we found that most acted in the expected direction (Figure 1D). For all five TFs, <0.9% of bQTLs were located in predicted motifs, and <1.3% of SNPs in predicted motifs were bQTLs (Figure S1D). This minor role of motifs suggests that many bQTLs may involve recruitment of one TF by another, which can occur in the absence of a binding motif for the recruited TF; and/or that sequences flanking the binding motif play a major role, as has been observed before (White et al., 2013; Dror et al., 2015; Levo et al., 2015).

To investigate the bQTLs falling outside motifs, we tested the idea that the GC content surrounding motifs plays an important role in TF binding (White et al., 2013; Dror et al., 2015). Specifically, it has been observed that high local GC content promotes the binding of some TFs, whereas high AT content is ideal for others (Dror et al., 2015). To test this in our data, we asked whether the high-affinity bQTL alleles showed a skewed GC content, particularly when close to a binding motif. For NF-κB bQTLs, we observed a significant enrichment for high-affinity G/C alleles specifically within 25 bp of the motif (Fisher's exact $p = 0.005$; Figure S1E). In contrast, Stat1 showed the opposite pattern, with a strong preference for high-affinity A/T alleles near the motif (Fisher's exact $p = 0.003$). None of the bQTLs showed any significant GC content bias at distances further than 25 bp from the motif (Figure S1E). Therefore, our results support the idea that for some TFs, GC content surrounding binding motifs can play an important role in determining affinity.

Comparing Pooled bQTLs with ChIP-Seq in Individual Samples

To validate our pooled QTL mapping approach, we asked how well our bQTLs can predict allele-specific binding in individual LCLs heterozygous for those bQTLs. Specifically, we compared our PU.1 bQTLs with PU.1 ChIP-seq data from 47 individual LCLs (Waszak et al., 2015); these samples were from another population (European/CEU), so were independent of those used in our experiments. We found that out of 592 of our PU.1 bQTLs (covered by at least 20 reads in the 47 CEU samples when summing across all heterozygous individuals), 518 (88%) showed an allelic bias in the direction predicted by our bQTLs, compared to 50% expected by chance (Figure 2A). However, since the allelic bias for most of these was of a small magnitude and not statistically significant, we then asked how well our bQTLs predict the individual-level allele-specific binding when we restrict the comparison to the 998 PU.1 bQTLs reported by Waszak et al. (2015). In this case, the agreement is much stronger (88/89 SNPs, 99%; Figure 2B), and notably the one discordant SNP was also the one with the fewest reads (summed across both datasets).

As a second test of validation, we asked whether the quantitative effect sizes inferred in our pooled data were also in agreement with the strength of allele-specific binding in individual heterozygotes. Among the 88 bQTLs identified above, the effect sizes were well correlated (Pearson's $r = 0.575$; $p = 5 \times 10^{-9}$), and the agreement was higher still for those bQTLs covered by more reads (e.g., for the 61 bQTLs with at least 100 reads in our data, Pearson's $r = 0.705$; $p = 2 \times 10^{-10}$; Figure 2C). Altogether, these results demonstrate a high degree of concordance between our pooling-based bQTLs and individual-level data.

Another important question is how the cost and effort of our pooled-QTL approach compares with standard QTL mapping. To investigate this, we again compared our PU.1 bQTLs with those previously mapped in 47 individual LCLs (Waszak et al., 2015). To ensure a fair comparison, we sought a p value cutoff for our PU.1 bQTLs that yielded similar enrichments for an independent set of QTLs, suggesting that the two lists are of comparable quality. We therefore intersected both bQTL lists with published QTLs for three histone modifications (hQTLs) (Grubert et al., 2015). At a $p < 5 \times 10^{-5}$ cutoff for our bQTLs, we found comparable levels of enrichment for the two PU.1 bQTL lists (Figure S2; slightly higher enrichments for our bQTLs in two out of three comparisons). At this cutoff, we mapped 3,246 bQTLs, compared to 998 for the standard, non-pooled approach. We achieved this despite generating 7.6-fold fewer ChIP-seq reads and performing 23.5-fold fewer ChIPs (Figure 3D). Therefore, in terms of cost per QTL, we estimate that for PU.1 our method is over 25-fold more effective than the standard approach. This difference will likely increase with the number of samples studied, since pooling provides the greatest savings for large sample sizes.

We hypothesize that these improvements are due to two key aspects of our approach. First, by performing ChIP with all samples pooled together, we reduce the inevitable experimental variability between samples. Second, our approach takes full advantage of the allele-specific information in each read, whereas traditional QTL mapping—by focusing on the mean of

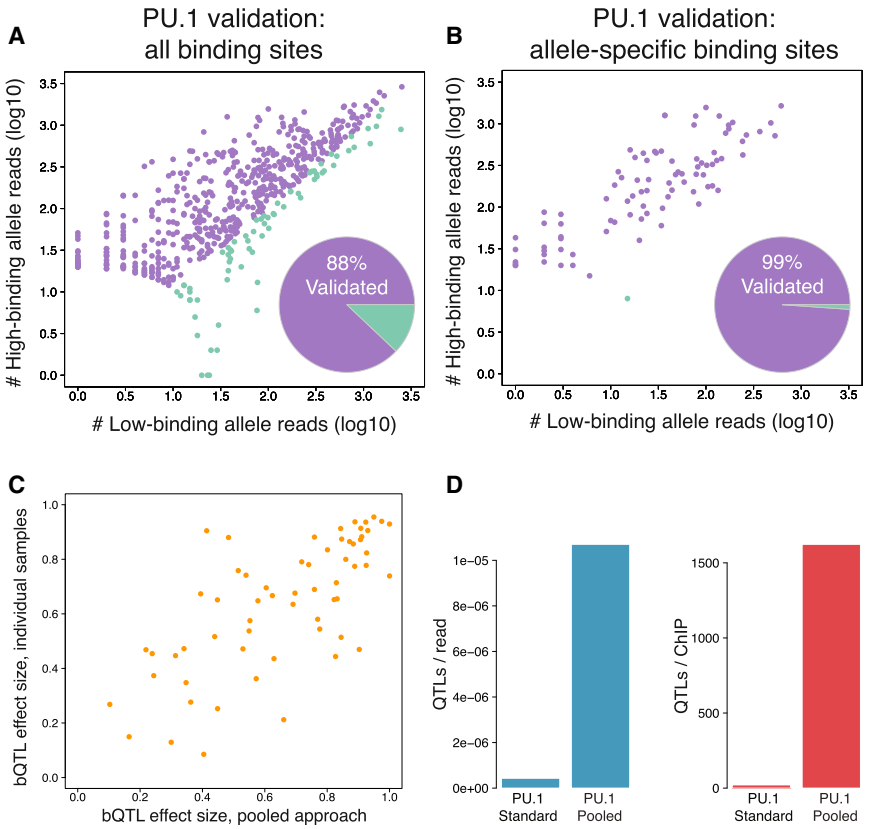


Figure 2. Comparing Pooled bQTLs with ChIP-Seq in Individual Samples

(A) Examining allele-specific binding of PU.1 (measured in individual heterozygous samples) at 592 PU.1 bQTLs (Waszak et al., 2015), we found 88% directionality agreement.

(B) Comparing our PU.1 bQTLs to allele-specific binding at 89 SNPs previously reported as PU.1 bQTLs (Waszak et al., 2015), we found 99% directionality agreement.

(C) Effect sizes of our bQTLs compared to the strength of allelic bias summed across individual heterozygous samples.

(D) Number of PU.1 bQTLs per sequence read and per ChIP, at equivalent significance cutoffs for each dataset.

See also Figure S2.

both alleles in each sample—ignores an important source of information in heterozygous samples (Kaplow et al., 2015; van de Geijn et al., 2015). In the extreme case of a cohort with only heterozygotes at a particular SNP, traditional QTL mapping would have no power, in contrast to our pooling approach. However, we note that pooling is not necessary to utilize this extra information, since it is also present in individual heterozygotes (van de Geijn et al., 2015).

Comparing bQTLs with Other Molecular QTLs

Many other types of molecular-level QTLs have been mapped in LCLs, including QTLs for mRNA levels (eQTLs), chromatin accessibility (dsQTLs), DNA methylation (mQTLs), and histone modifications (Fraser and Xie, 2009; Degner et al., 2012; Fraser et al., 2012; Kasowski et al., 2013; Lappalainen et al., 2013; McVicker et al., 2013; Kilpinen et al., 2013; Gutierrez-Arcelus et al., 2013; Ding et al., 2014; Banovich et al., 2014; Waszak et al., 2015; Grubert et al., 2015). To explore the relationships between TF binding and these other regulatory levels, we intersected our bQTLs with each of these other QTL types and calculated the fold-enrichment compared to SNPs bound by our TFs but not classified as bQTLs (to control for any general characteristics of each TF's binding sites, such as proximity to promoter regions) (see the Supplemental Experimental Procedures). We found dsQTLs to have the highest overall enrichment for bQTLs (9.2-fold median enrichment; Figure S3A) as expected, since these are the most directly linked to TF binding (Degner et al., 2012). mQTLs had the lowest levels of enrichment (Figure S3A), suggesting that binding of these TFs is largely independent of DNA methylation.

Because our TFs are all primarily transcriptional activators, we expected that bQTLs would be biased in their direction of effects on other molecular traits—alleles with higher binding should typically have more accessible chromatin, greater local H3K4me3, and higher mRNA levels. To investigate this, we calculated the directionality agreement between bQTLs and these other QTL types, for SNPs found as QTLs for both. In each case, we found increasing agreement at more stringent bQTL p values (Figures 3 and S3B). For previously published H3K4me3 QTLs (Grubert et al., 2015), this reached 93.1%–96.1% agreement for four out of five TFs and 94.9% agreement with our H3K4me3 QTLs. Since false positive bQTLs would agree only ~50% of the time, this suggests false discovery rates of between 7.8%–13.8% for these bQTLs (although these estimates, which conservatively assume that the previously reported H3K4me3 QTLs have no false positives and that the true overlaps always agree in direction, cannot be extrapolated to all bQTLs) (see the Supplemental Experimental Procedures). The one TF with lower (85.9%) agreement, Pou2f1, has been shown to sometimes act as a repressor (Lin et al., 2013), which may explain this difference. We observed almost identical concordance among bQTLs occurring outside of their TF's consensus motifs (Figure S3C), suggesting that these bQTLs are not enriched for false positives.

Finally, we also compared our bQTLs with SNPs that themselves affect transcription (in contrast to the QTL comparisons above, where linkage disequilibrium generally prevents identification of causal variants). These SNPs, identified in a massively parallel reporter assay performed in LCLs (Tewhey et al., 2016), showed 79.2%–100% agreement in directionality with our bQTLs for all TFs (Figure 3D). Altogether, the concordance of bQTLs with four other types of molecular QTLs (Figure 3) further establishes the accuracy of our approach.

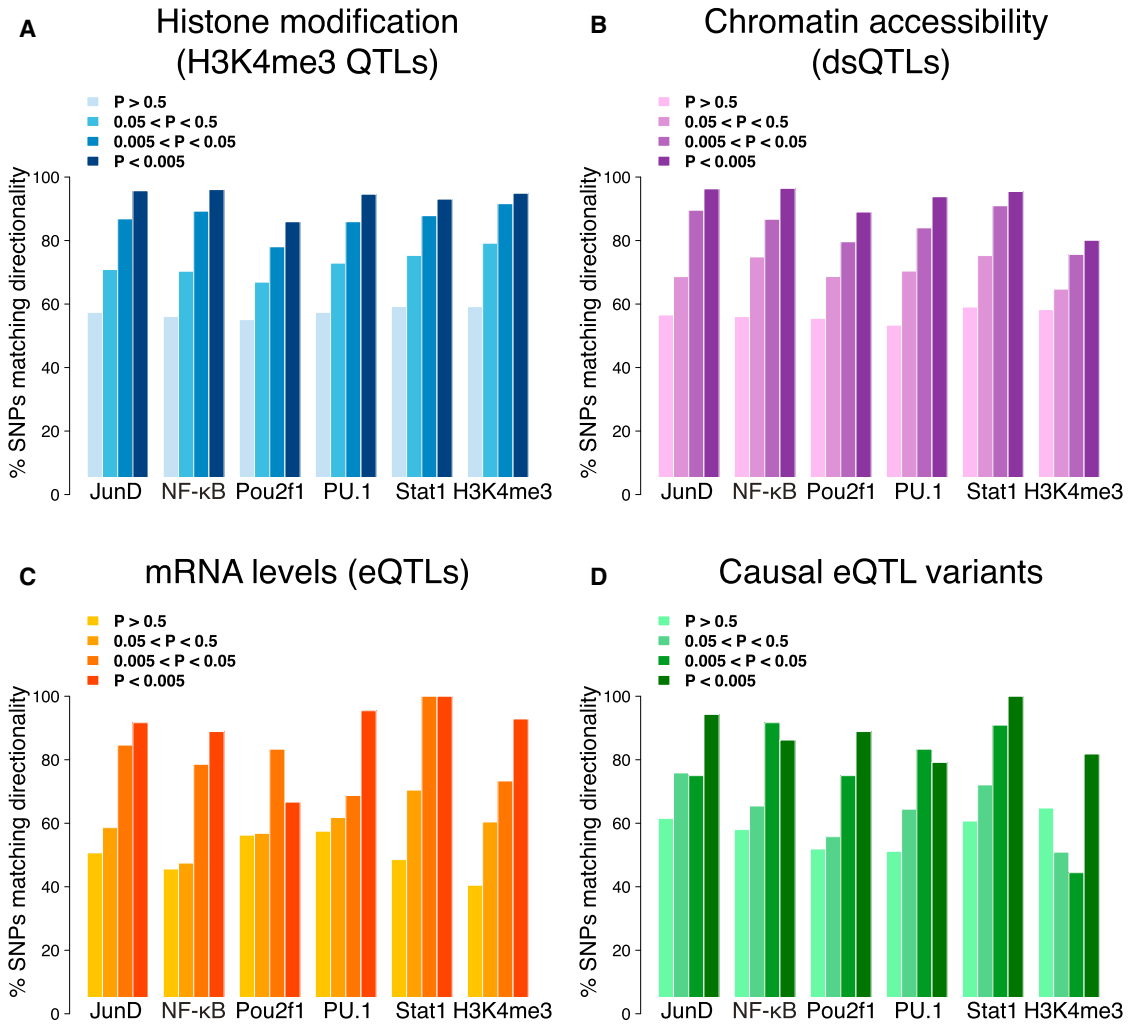


Figure 3. Comparisons with Other Molecular-Level QTLs

For each type of previously published QTL, we calculated the degree of allelic concordance with our bQTLs.

(A) H3K4me3 QTLs (Grubert et al., 2015).

(B) dsQTLs (Degner et al., 2012).

(C) eQTLs (Lappalainen et al., 2013).

(D) Causal SNPs underlying eQTLs (Tewhey et al., 2016).

See also Figure S3.

Inferring Cooperative Binding and Pioneer Factors

Our data allow us to test whether a bQTL for one TF is likely to act as a bQTL for other TFs as well. For example, rs2886870—our second strongest bQTL for NF-κB ($p = 10^{-75}$)—was also a bQTL for JunD ($p = 10^{-20}$) and PU.1 ($p = 10^{-16}$). It disrupts an NF-κB motif, suggesting that its direct effect is on NF-κB binding, which then affects recruitment of additional TFs. Interestingly, this SNP has previously been associated with local chromatin accessibility, histone modifications, RNA polymerase II binding, and mRNA levels in LCLs (McVicker et al., 2013). It is also the strongest QTL for DNA replication timing in LCLs ($p = 10^{-13}$) (Koren et al., 2014), affecting an ~422 kb region, suggesting that disrupting a single NF-κB binding site that recruits multiple TFs can have a major effect on the temporal program of DNA replication, in addition to local chromatin and transcription.

Examining the overlap between bQTLs for all of our factors, we found a 2.7-fold to 21-fold enrichment, with Stat1 showing the strongest overall enrichment (Figure 4A, upper right). This suggests that single genetic variants may often affect binding of multiple TFs, since the alternative—multiple causal variants for different TFs in strong LD—is unlikely to occur so consistently. To further test this idea, we measured the concordance in directionality for bQTLs associated with binding of multiple TFs, which would be random (~50%) if one bQTL was tagging multiple causal variants. We found that the same allele had higher binding for both TFs in 96.1%–100% of all TF pairs (Figure 4A, lower left), strongly supporting the single-variant model.

A single variant might affect multiple TFs by initially altering the binding of one TF—a "pioneer factor"—that in turn recruits others. To test this idea, we investigated bQTLs for one TF that

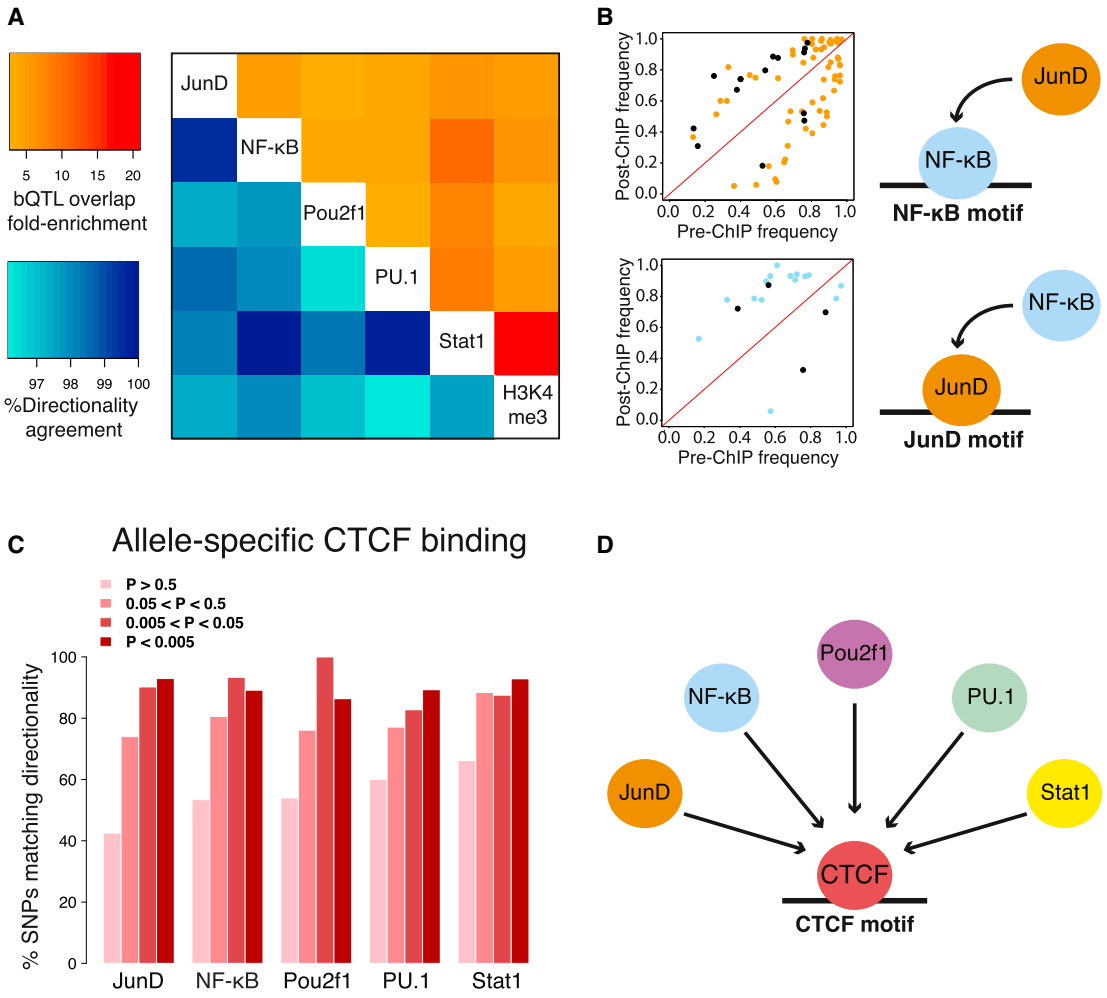


Figure 4. Causal Relationships between TFs

(A) For every pairwise combination of bQTLs, we calculated the enrichment of their overlaps (upper right, relative to bound SNPs that are not bQTLs) and the degree of allelic concordance among the overlaps (lower left).

(B) We tested whether SNPs in the binding motifs for one TF were predictive of binding of another TF. Scatter plots show allele frequencies before and after ChIP (as in Figure 1B), with those SNPs whose effect is in the direction predicted by the motif's PWM colored to match the ChIP-ed TF and those not matching colored black.

(C) Allelic bias of CTCF binding at heterozygous sites (Ding et al., 2014) generally agrees in direction with our bQTLs. See also Figure S4.

(D) Our model for CTCF as a major pioneer factor.

are located within a binding motif for a different TF, since these suggest that one TF recruits another to its binding site (Kasowski et al., 2010)—as in the case of rs2886870 and NF-κB described above (this causality inference is only possible when a bQTL is within a TF binding motif, since otherwise we cannot discriminate between direct versus indirect effects of a bQTL on TF binding). For example, NF-κB and JunD are known to act cooperatively (Rahmani et al., 2001); consistent with this, we found that SNPs in either of their motifs often associate with binding of the other, implying that either factor can help recruit the other (Figure 4B). In further support of this causal inference approach, the directionality of these shared bQTLs is overwhelmingly in the direction predicted by the motif's PWM: for example, a SNP increasing the predicted affinity of a site for NF-κB usually increases the binding of JunD (Figure 4B, top, orange points) rather than decreasing it (Figure 4B, top, black points).

Extending this test to a collection of 2,995 motifs (Weirauch et al., 2014), we found one TF whose motif disruption was associated with binding of all five TFs (Fisher's exact $p < 0.05$ for each). This motif was for CTCF, a protein involved in chromosomal looping and chromatin structure (Rao et al., 2014). To further explore this relationship, we tested the directionality agreement between our bQTLs and QTLs for CTCF binding in LCLs (Ding et al., 2014). These showed 86.4%–92.9% agreement for all five TFs (Figures 4C and S4), supporting the motif analysis. Together, these results suggest that CTCF may be a major pioneer factor that influences the local landscape of binding for many other TFs (Figure 4D), although our results do not reveal the mechanism by which this occurs. It is possible that

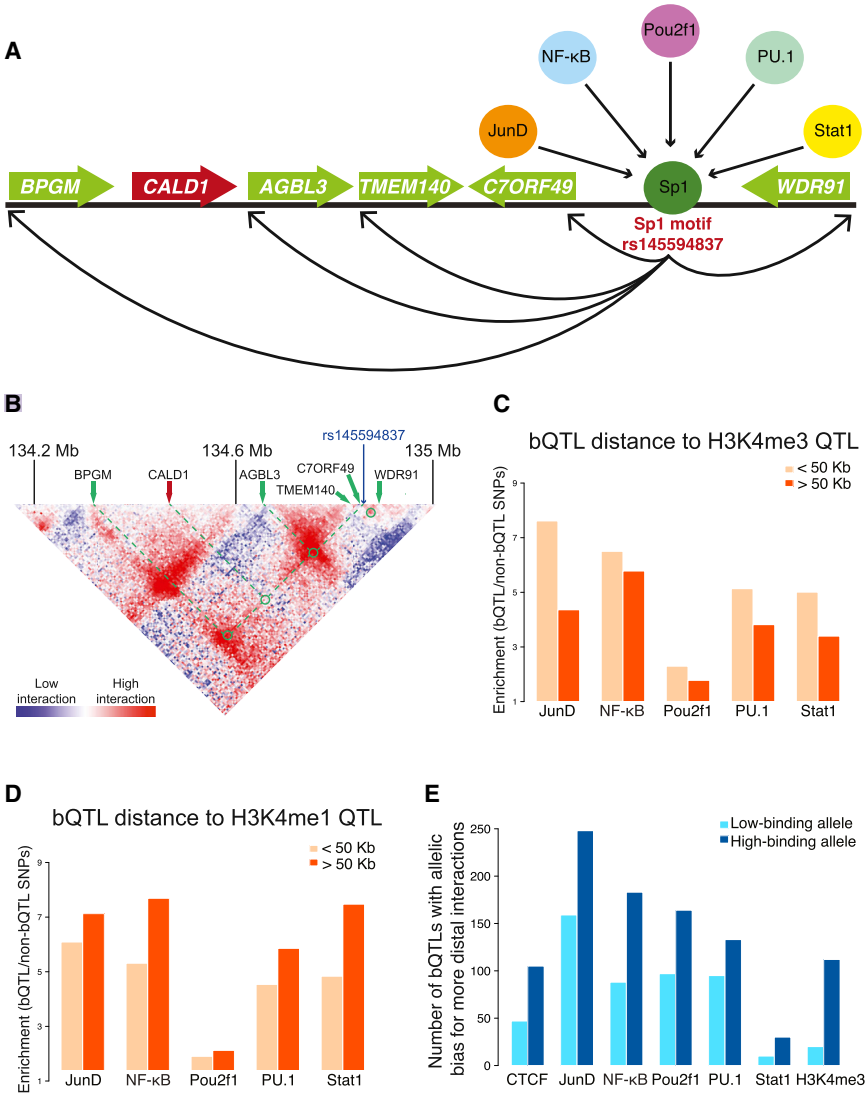


Figure 5. Long-Range Effects of bQTLs

(A) An example of a bQTL for all five TFs that affects the expression of five genes, including one (*BPGM*) 525 kb away.

(B) The pattern of chromosomal contacts around the bQTL shown above indicates high levels of interactions with the promoter regions of the distal genes that it affects. No enrichment is observed with *CALD1*, a gene within this region not affected by the bQTL. Color bar indicates the ratio of observed/expected contact, given the distance between loci (even though the deletion is not enriched more than expected for contacts with the two closest genes, *C7ORF49* and *TMEM140*, there is still a high level of contact with these and other proximal loci). Hi-C data were normalized with the "balanced" method (Rao et al., 2014). See also Figure S5.

(C) Enrichment of bQTLs for local and distal hQTLs affecting H3K4me3 show a consistent bias for local effects.

(D) Enrichment of bQTLs for local and distal hQTLs affecting H3K4me1 show a consistent bias for distal effects.

(E) bQTLs affect the extent of long-range contacts. For each factor, the number of bQTLs where the higher-bound allele has more distal (>30 kb) interactions is plotted in dark blue, and the number where the lower-bound allele has more is plotted in light blue. The difference is significant in every case.

there are direct physical interactions between CTCF and the other TFs, but a more likely model is that CTCF binding alters nucleosome positioning and chromosomal looping to make local chromatin more accessible to other TFs.

bQTLs Can Mediate Long-Range Effects on Transcription

Cis-regulatory variants typically affect chromatin and transcription in a highly local manner—for example, most dsQTLs are located within 100 bp of the region whose accessibility they modulate (Degner et al., 2012). However, long-range effects of *cis*-regulatory variants are also possible, particularly when distant chromosomal regions are in close physical proximity within the nucleus (Grubert et al., 2015).

To investigate whether bQTLs can affect distant loci, we first compared them with long-range eQTLs. One example of a bQTL with distant effects is rs145594837, a 10-bp deletion that creates a strong Sp1 motif within an enhancer/promoter that is active in dozens of cell types (ENCODE Project Consortium, 2012; Kundaje et al., 2015). The deletion allele has higher binding of all five TFs, suggesting that Sp1 may recruit them to this locus when it is bound. This deletion is an eQTL for five genes (Lappalainen et al., 2013), including *BPGM*, a gene involved in congenital erythrocytosis that is located 525 kb away (Figure 5A). Interestingly, the deletion is not an eQTL for a more proximal gene, *CALD1*, that is adjacent to *BPGM*. To determine if chromosomal contacts might mediate these selective effects on transcription, we analyzed Hi-C data from an LCL (Rao et al., 2014), which indicate which loci are in close proximity within the nucleus. Examining interactions between the deletion locus and each gene's promoter, we found that all five genes affected by the deletion had high rates of contact, while *CALD1* had the lowest level of interaction with the deletion locus (Figure 5B). This pattern suggests that this bQTL likely affects distant genes via long-range chromosomal contacts.

To more systematically explore the long-range effects of bQTLs, we compared them to QTLs previously found to affect histone modifications (hQTLs); these hQTLs were previously classified as either local (<50 kb from the site they affect) or distal (>50 kb away) (Grubert et al., 2015). For hQTLs affecting the promoter-associated mark H3K4me3, we found a greater enrichment for bQTLs acting as local as opposed to distal hQTLs

(Figure 5C). Although the magnitude of the difference was small, it was consistent across all five TFs (Fisher's exact $p = 3.9 \times 0^{-7}$), indicating that short-range effects on H3K4me3 are more likely to be explained by bQTLs. In contrast, for hQTLs affecting the enhancer-associated mark H3K4me1, we found the opposite trend, with a slightly higher enrichment of bQTLs with long-range effects for all five TFs (Figure 5D; Fisher's exact $p = 1.1 \times 10^{-8}$). These opposing patterns may be due to enhancers generally participating in more long-range chromosomal contacts than promoters, giving more opportunity for bQTLs to act on enhancers from a distance. Consistent with this model, we found no distance bias for hQTLs affecting H3K27ac (Fisher's exact $p = 0.27$), a mark found at both enhancers and promoters. Across all hQTLs, the most distant effect we observed was a PU.1 bQTL (rs10161851) associated with H3K4me1 levels 1.98 Mb away; however, since these hQTLs were only mapped within 2 Mb (Grubert et al., 2015), it is likely there are bQTLs affecting chromatin across even greater distances.

Another question is whether these long-range chromosomal contacts act simply as passive conduits for bQTLs affecting distal loci, or if instead bQTLs can help shape the three-dimensional architecture of chromosomes and thus affect which other loci they contact. CTCF has been found at the base of many chromosomal loops (Rao et al., 2014), and altering these sites can disrupt looping (Guo et al., 2015b; Sanborn et al., 2015), but the role of natural genetic variants in chromosomal architecture is unknown. To investigate this, we tested whether bQTLs for CTCF or other TFs may affect the extent of long-range contacts at these loci. Utilizing allele-specific Hi-C data from an LCL (Rao et al., 2014), for each bQTL that was heterozygous in this sample we counted the number of Hi-C reads connecting each allele of the bQTL with any distal locus (>30 kb away). For each TF, we then counted the number of bQTLs where the high-binding allele had significantly more contacts (based on a binomial test of allele-specific read counts) and likewise counted the cases where the low-binding allele was more interactive. For previously published CTCF bQTLs (Ding et al., 2014), we found that high-binding alleles were 2.2-fold more likely to have more contacts than the low-binding alleles (Figure 5E; binomial $p = 1.4 \times 10^{-6}$), consistent with CTCF's well-established role in shaping chromosomal contacts. Surprisingly, we observed a similar pattern for all five of our TFs (Figure 5E; 1.4-fold to 3.0-fold differences; binomial $p \leq 0.001$ for each). At sites where multiple heterozygous bQTLs were present and their high-binding alleles were encoded on the same chromosome, the trends were stronger still (3.0-fold to 5.0-fold differences for the five TFs and 6.0-fold for CTCF; binomial $p \leq 0.001$ for each). We found the strongest effects of all for H3K4me3 bQTLs, with a 5.6-fold interaction difference between alleles with more versus less of this modification (binomial $p = 5.2 \times 10^{-17}$) and 9.5-fold when these were accompanied by a nearby TF bQTL (binomial $p = 8.2 \times 10^{-12}$). Although our bQTLs are also often associated with CTCF binding (Figure 4C), CTCF could not explain these patterns at our bQTLs, since excluding sites with detectable CTCF binding did not weaken the overall trend (1.6-fold to 2.7-fold differences for four out of five TFs and 12-fold for H3K4me3). Altogether, these results suggest that natural genetic variation can affect chromosomal architecture, and the alleles with greater TF binding or more H3K4me3 tend to have more distal contacts, even at loci lacking CTCF.

bQTLs Provide Candidate Molecular Mechanisms for Thousands of GWAS-Implicated Variants

Integrating our results with GWAS can help elucidate molecular mechanisms of complex disease. For example, rs10263017—a bQTL for NF-κB whose derived allele disrupts an NF-κB binding motif (circled in Figure 1B)—is also associated with mRNA levels of *EIF2AK1* (Lappalainen et al., 2013), as well as risk of myocardial infarction (Yamada et al., 2011). Thus, our results imply that its likely molecular mechanism is alteration of NF-κB binding (Figure 6A). This variant is located in an enhancer that is active in dozens of tissues, including the heart left ventricle (Kundaje et al., 2015), illustrating how bQTLs from LCLs can reveal molecular mechanisms that may be most relevant in other cell types.

In other cases, bQTLs affecting disease risk may act over long genomic distances. One example of this is rs7122786, a SNP located in an LD block that contains multiple independent variants associated with risk of prostate cancer (Amin Al Olama et al., 2015). The derived allele disrupts an NF-κB binding motif within an LCL-specific DNase hypersensitive region (ENCODE Project Consortium, 2012) and is a bQTL for NF-κB. Although it is located near the *MYEOV* gene, it is actually an eQTL for another gene 427 kb away: *CCND1* (cyclin D1), a cell-cycle regulator that plays a major role in prostate cancer and other malignancies. This variant was the strongest eQTL for *CCND1* in LCLs ($p = 3 \times 10^{-10}$) (Dixon et al., 2007) and was replicated in another study (Lappalainen et al., 2013), further supporting the hypothesis that this is a functional variant. Interestingly, two other enhancers located 125 kb and 216 kb from *CCND1* each contain variants associated with risk of specific cancer types (ER^+ breast cancer and renal cell carcinoma, respectively) (Schödel et al., 2012; French et al., 2013); our results suggest the existence of a third enhancer that specifically affects risk of prostate cancer (Figure 6B). Hi-C data (Rao et al., 2014) show that all three enhancers and *CCND1* are part of a large region with a high density of chromosomal interactions (Figure S5), which likely facilitates this long-range effect.

We can also infer more complex molecular mechanisms involving TFs recruited to the binding sites of other factors. For example, variants in two LD blocks flanking *IRGM*, a regulator of autophagy, are independently associated with Crohn's disease (Parkes et al., 2007; McCarroll et al., 2008). For one of these blocks, two candidate causal variants have been proposed (McCarroll et al., 2008; Brest et al., 2011); however, in the second block, the strong LD and lack of obvious candidates has prevented any progress in determining its causal variant. We found that rs6893009, which is in perfect LD with the reported variant (rs4958847), is one of our strongest bQTLs for PU.1 ($p = 2 \times 10^{-31}$), as well as a moderate bQTL for NF-κB and JunD ($p < 2 \times 10^{-5}$). The derived allele creates a high-affinity PU.1 binding site, suggesting that PU.1 helps recruit NF-κB and JunD (Figure 6C). In addition to B cells and LCLs, this region has an active chromatin state in the small intestine (Kundaje et al., 2015), a key tissue for Crohn's disease. A second example involves rs194749, a bQTL for JunD and PU.1 that is also associated

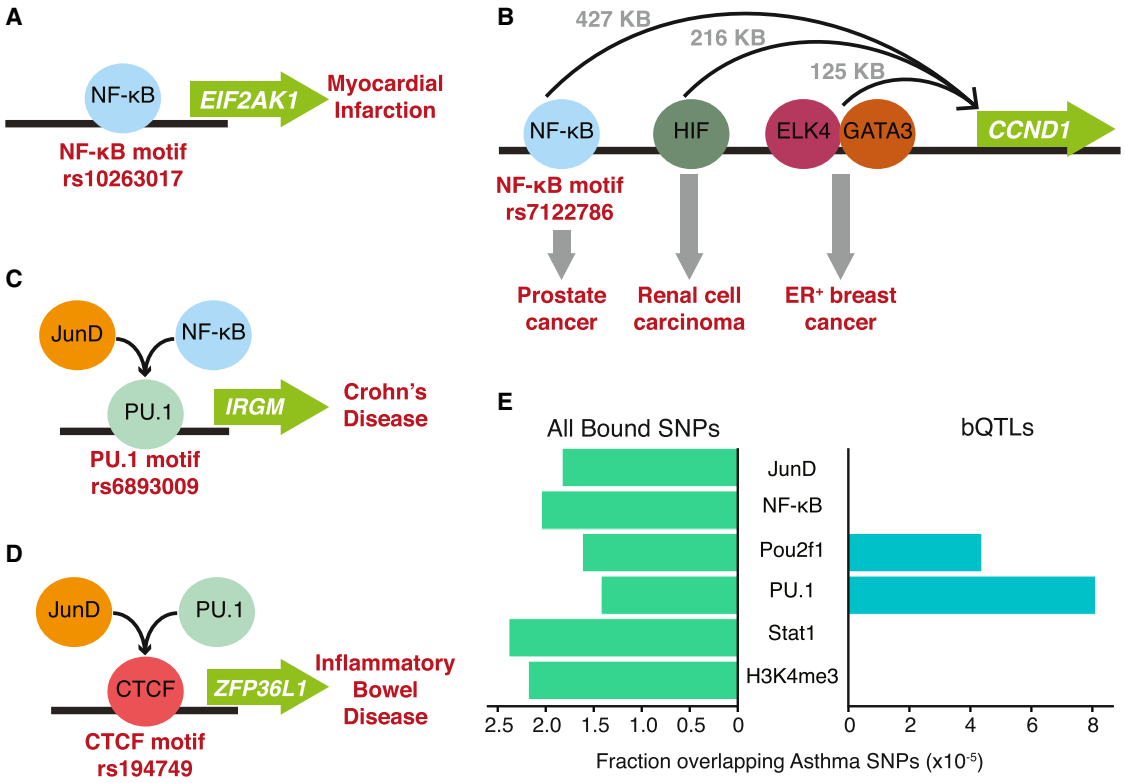


Figure 6. Effects of bQTLs on Disease Risk

Each example involves a SNP in a TF binding motif that is also a bQTL and is associated with disease risk.

(A) bQTL for NF-κB (highlighted in Figure 1B) in an NF-κB binding motif that is also an eQTL for *EIF2AK1* and associated with myocardial infarction.

(B) bQTL for NF-κB in an NF-κB binding motif that is also an eQTL for *CCND1* and is associated with prostate cancer. See also Figure S5.

(C) bQTL for PU.1, NF-κB, and JunD in a PU.1 binding motif that is also associated with Crohn's disease.

(D) bQTL for JunD and PU.1 in a CTCF binding motif that is also associated with inflammatory bowel disease.

(E) Comparing enrichments for asthma-associated SNPs: all bound SNPs (left) and bQTLs (right).

See also Figure S6 and Table S2.

with inflammatory bowel disease (Jostins et al., 2012). This SNP falls within a CTCF binding motif and is also associated with local chromatin accessibility (Degner et al., 2012) and DNA methylation (Gutierrez-Arcelus et al., 2013); the nearest gene is *ZFP36L1*, which has a well-established anti-inflammatory role (Sanduja et al., 2011). Therefore, it is likely that CTCF is the pioneer factor, whose binding impacts the recruitment of JunD and PU.1, potentially altering the *cis*-regulation of *ZFP36L1* (Figure 6D).

Examining disease-specific bQTL/GWAS enrichment also revealed informative trends. In these comparisons, we can go beyond comparing GWAS loci with all binding sites of a TF (Schaub et al., 2012; Farh et al., 2015; Guo et al., 2015a), since most SNPs in binding sites do not detectably affect binding (Figure S1D). For example, asthma risk loci were enriched only for bQTLs of PU.1 and Pou2f1 (5.7-fold and 2.7-fold enriched, respectively; Figure 6E right). However across all binding sites (not just bQTLs), PU.1 and Pou2f1 actually had the fewest overlaps with asthma-associated SNPs (Figure 6E, left). We found a similar lack of correspondence for many other diseases (median correlation of bQTL/GWAS overlap versus bound SNP/GWAS overlap of $r = -0.03$ among 12 autoimmune diseases; Figure S6), suggesting that intersecting GWAS loci with all binding sites of a TF may yield misleading results because these overlaps are dominated by SNPs with no effect on TF binding.

Altogether, 3,601 of our bQTLs have been previously implicated by GWAS, either directly or indirectly via LD ($r^2 > 0.8$ in YRI) (Table S2). These represent 2,282 different disease-associated variants, 995 (44%) of which are associated with bQTLs for multiple factors. Interestingly, this is 8.0-fold higher than the overall fraction of bQTLs associated with multiple factors, suggesting that variants affecting multiple TFs are more likely to impact phenotypes. Certain TF combinations may be especially likely to affect particular traits: for example, although PU.1 and Stat1 are known to act cooperatively (Aittomäki et al., 2004), they have few shared bQTLs. Nonetheless, two of their shared bQTLs are associated with atopic dermatitis (Sun et al., 2011; Hirota et al., 2012)—over 5,000-fold more than expected by chance ($p = 7 \times 10^{-8}$). No other disease shows a significant enrichment for this TF pair, suggesting the possibility of a specific role for these TFs in atopic dermatitis, although additional examples will be required to confirm this trend.

DISCUSSION

In this work, we have introduced an approach for mapping *cis*-acting QTLs for molecular traits. Compared to the standard QTL mapping approach, we achieved a >25-fold reduction in cost and effort. We applied this method to map thousands of QTLs for TF binding and histone methylation, revealing a widespread role for common genetic variants to modulate TF binding in humans. We found that single genetic variants can often affect the binding of multiple TFs; that CTCF is a major pioneer factor capable of recruiting diverse TFs to its binding sites; that bQTLs can have long-range effects on *cis*-regulation, mediated by chromosomal contacts; and that bQTLs can also affect the extent of these long-range contacts.

Our collection of bQTLs is also a valuable resource for any researchers investigating the molecular mechanisms of disease risk loci. Indeed, our observation that disease-specific bQTL enrichments differ substantially from simple TF binding site enrichments (Figure 6E) suggests that bQTLs provide a critical level of information regarding disease etiologies. However, since any given bQTL/GWAS variant overlap could be due to chance, these should be regarded as mechanistic hypotheses to guide further investigation.

Although it is clear that the vast majority of GWAS-implicated variants are noncoding and many fall into tissue-specific enhancers, their effects on transcription or other cellular processes are still mostly unknown. Considering the extensive GWAS overlap we found with bQTLs for only five TFs in one cell type, our work suggests that a substantial fraction of genetic variation affecting human phenotypes may act via TF binding. Indeed, the number of GWAS hits in our bQTL catalog is greater than has been reported in recent eQTL studies analyzing RNA-seq data from hundreds or thousands of samples (Lappalainen et al., 2013; GTEx Consortium, 2015). How could these bQTLs affect phenotypes, if they are not also eQTLs? Although some bQTLs may act through processes other than transcription (e.g., replication timing), a more likely explanation is that their effects on transcription have not been detectable thus far. This could be due to small effect sizes, tissue-specificity, or environmental-specificity; or alternatively, some may only affect transcriptional variability, either between cells or between individuals (Ansel et al., 2008; Fraser and Schadt, 2010; Hulse and Cai, 2013). Of course, many bQTLs may have no effects on transcription or downstream phenotypes, although this will be difficult to distinguish from other scenarios listed above.

Our pooled ChIP-seq approach can be extended in many ways. Larger pools will allow inclusion of low-frequency variants, and deeper sequencing will allow detection of smaller effects. Our approach can be applied to any cell type from any species, with only small amounts of biological material needed from each individual. In addition, our pooling framework is not limited to ChIP—any method that isolates a subset of DNA or RNA could be used in its place. For example, our approach could map variants in DNA affecting chromatin accessibility or noncoding RNA localization, or variants in RNAs affecting RNA-protein interactions or translational pausing. Other pooling-based frameworks that do not involve isolating a subset of DNA/RNA are also possible, as we have shown for mapping QTLs affecting DNA methylation (Kaplow et al., 2015). Large collections of *cis*-regulatory variants, such as those we present here, may ultimately enable the development of predictive models of TF binding and transcription.

EXPERIMENTAL PROCEDURES

Cell Culture Conditions

LCLs from unrelated Yoruban individuals were obtained from the Coriell Institute. The LCLs were grown in RPMI-GlutaMAX-HEPES media (Life Technologies) with 15% FBS, 100 IU/ml penicillin, and 100 μg/ml streptomycin at 37°C, 5% CO_2.

ChIP-Seq

Each individual cell line was grown to a density of 6–8 × 10^5 cells/ml and 7 × 10^5 cells were collected. Samples were pooled such that each individual was approximately equally represented in the pool. TF ChIPs were done using a pool of 60 individuals. The cells used for NF-κB ChIP were first treated with 25 ng/ml human recombinant TNF-alpha (eBioscience #14-8329) for 6 hr at 37°C, 5% CO_2. After stimulation, cells were cross-linked in 1% formaldehyde for 10 min at room temperature. Nuclear lysates were sonicated using a Branson 250 Sonifier (power setting 2, 100% duty cycle for 7 × 30-s intervals), yielding chromatin fragments of 50–2,000 bp. Lysates were treated overnight at 4°C with 8 μg of antibody (Santa Cruz Biotechnology: JunD [sc-74], NF-κB [sc-372], Pou2f1/Oct-1 [sc-232], PU.1 [sc-22805], Stat1 [sc-345]). Protein-DNA complexes were captured on Dynabeads-Protein G (Life Technologies, #10003D) and eluted in 1% SDS TE buffer at 65°C. Purified ChIP-seq libraries were generated as described in Kasowski et al. (2010), with two biological replicates per TF, and sequenced on an Illumina HiSeq 2000 (101 bp, paired-end reads). H3K4me3 ChIP was performed as described in McVicker et al. (2013), using 4 μg of antibody for H3K4me3 (Abcam ab8580) in a pool of 71 LCLs. Libraries were sequenced on an Illumina HiSeq 2000 (50 bp, single-end reads). All data are available at NCBI SRA (SRP058204).

Calculating Post-ChIP Allele Frequencies

Genotypes were imputed for all individuals for 18,931,934 SNPs, using a published imputation pipeline (Degner et al., 2012). Reads were mapped using the WASP pipeline (van de Geijn et al., 2015). Briefly, raw fastq reads were mapped to the human reference sequence (GRCh37) using bowtie2 (with parameters: -N 0 -L 20–end-to-end–np 0–n-ceil 0,0.15) (Langmead and Salzberg, 2012). We used the "find_intersecting_snps.py" script from the WASP pipeline to identify reads that have mapping biases, remapped with the same bowtie2 parameters, and used "filter_remapped_reads.py" to retrieve reads that remapped correctly. These mapped BAM files were sorted using samtools sort and duplicate reads were filtered out for each replicate using samtools rmdup (Li et al., 2009). After duplicate removal, replicates were combined for all subsequent analysis. Samtools mpileup was used to retrieve all mapped reads aligning to each SNP (with parameters: -Q 25 -f < hg19_genome > -l < SNPs.bed >).

Reads were filtered out if they had a base quality score <25 at the SNP. SNPs were excluded if they had a minor allele frequency (MAF) <0.02 among the 60 YRI samples, or fewer than 15 reads mapping to either allele, or if they were located in an ENCODE "blacklisted" region (ENCODE Project Consortium, 2012). The post-ChIP allele frequencies were calculated by counting the number of reads from each allele. Although post-ChIP frequencies could be inferred for unobserved SNPs that are in perfect LD with observed ones, this would not result in any additional independent bQTLs, so we restricted our analysis to directly observed SNPs.

Mapping and Analyzing bQTLs

Additional methods are described in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

The accession number for the data reported in this paper is NCBI SRA: SRP058204.

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.2016.03.041.

AUTHOR CONTRIBUTIONS

H.B.F. conceived the project. A.K.T. performed the experiments and analysis. M.M. assisted with ChIP-seq of H3K4me3. T.M., D.G., B.L.H., and H.B.F. assisted with analysis. A.K.T. and H.B.F. wrote the manuscript.

ACKNOWLEDGMENTS

We would like to thank B. van de Geijn, Y. Gilad, T. Kawli, G. McVicker, J. Pritchard, M. Snyder, D. Spacek, R. Tewhey, and members of the H.B.F. lab for sharing data, protocols, cell lines, and advice. This work was supported by NIH grant 1DP2OD008456-01. A.K.T. is an NSF predoctoral fellow and was supported by NIH training grant T32GM007276.

Received: October 9, 2015
Revised: December 30, 2015
Accepted: March 23, 2016
Published: April 14, 2016

REFERENCES

Aittomäki, S., Yang, J., Scott, E.W., Simon, M.C., and Silvennoinen, O. (2004). Molecular basis of Stat1 and PU.1 cooperation in cytokine-induced Fcgamma receptor I promoter activation. Int. Immunol. *16*, 265–274.

Amin Al Olama, A., Dadaev, T., Hazelett, D.J., Li, Q., Leongamornlert, D., Saunders, E.J., Stephens, S., Cieza-Borrella, C., Whitmore, I., Benlloch Garcia, S., et al. (2015). Multiple novel prostate cancer susceptibility signals identified by fine-mapping of known risk loci among Europeans. Hum. Mol. Genet. *24*, 5589–5602.

Ansel, J., Bottin, H., Rodriguez-Beltran, C., Damon, C., Nagarajan, M., Fehrmann, S., François, J., and Yvert, G. (2008). Cell-to-cell stochastic variation in gene expression is a complex genetic trait. PLoS Genet. *4*, e1000049.

Banovich, N.E., Lan, X., McVicker, G., van de Geijn, B., Degner, J.F., Blischak, J.D., Roux, J., Pritchard, J.K., and Gilad, Y. (2014). Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels. PLoS Genet. *10*, e1004663.

Bonhoure, N., Bounova, G., Bernasconi, D., Praz, V., Lammers, F., Canella, D., Willis, I.M., Herr, W., Hernandez, N., and Delorenzi, M.; CycliX Consortium (2014). Quantifying ChIP-seq data: a spiking method providing an internal reference for sample-to-sample normalization. Genome Res. *24*, 1157–1168.

Brest, P., Lapaquette, P., Souidi, M., Lebrigand, K., Cesaro, A., Vouret-Craviari, V., Mari, B., Barbry, P., Mosnier, J.F., Hébuterne, X., et al. (2011). A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. Nat. Genet. *43*, 242–245.

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.

Ding, Z., Ni, Y., Timmer, S.W., Lee, B.K., Battenhouse, A., Louzada, S., Yang, F., Dunham, I., Crawford, G.E., Lieb, J.D., et al. (2014). Quantitative genetics of CTCF binding reveal local sequence effects and different modes of X-chromosome association. PLoS Genet. *10*, e1004798.

Dixon, A.L., Liang, L., Moffatt, M.F., Chen, W., Heath, S., Wong, K.C., Taylor, J., Burnett, E., Gut, I., Farrall, M., et al. (2007). A genome-wide association study of global gene expression. Nat. Genet. *39*, 1202–1207.

Dror, I., Golan, T., Levy, C., Rohs, R., and Mandel-Gutfreund, Y. (2015). A widespread role of the motif environment in transcription factor binding across diverse protein families. Genome Res. *25*, 1268–1280.

Edwards, S.L., Beesley, J., French, J.D., and Dunning, A.M. (2013). Beyond GWASs: illuminating the dark road from association to function. Am. J. Hum. Genet. *93*, 779–797.

Enard, D., Messer, P.W., and Petrov, D.A. (2014). Genome-wide signals of positive selection in human evolution. Genome Res. *24*, 885–895.

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

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

Fraser, H.B. (2013). Gene expression drives local adaptation in humans. Genome Res. *23*, 1089–1096.

Fraser, H.B., and Schadt, E.E. (2010). The quantitative genetics of phenotypic robustness. PLoS ONE *5*, e8635.

Fraser, H.B., and Xie, X. (2009). Common polymorphic transcript variation in human disease. Genome Res. *19*, 567–575.

Fraser, H.B., Lam, L.L., Neumann, S.M., and Kobor, M.S. (2012). Population-specificity of human DNA methylation. Genome Biol. *13*, R8.

French, J.D., Ghoussaini, M., Edwards, S.L., Meyer, K.B., Michailidou, K., Ahmed, S., Khan, S., Maranian, M.J., O'Reilly, M., Hillman, K.M., et al.; GENICA Network; kConFab Investigators (2013). Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers. Am. J. Hum. Genet. *92*, 489–503.

Grubert, F., Zaugg, J.B., Kasowski, M., Ursu, O., Spacek, D.V., Martin, A.R., Greenside, P., Srivas, R., Phanstiel, D.H., Pekowska, A., et al. (2015). Genetic control of chromatin states in humans involves local and distal chromosomal interactions. Cell *162*, 1051–1065.

GTEx Consortium (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science *348*, 648–660.

Guo, Y., Conti, D.V., and Wang, K. (2015a). Enlight: web-based integration of GWAS results with biological annotations. Bioinformatics *31*, 275–276.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., et al. (2015b). CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. Cell *162*, 900–910.

Gusev, A., Lee, S.H., Trynka, G., Finucane, H., Vilhjálmsson, B.J., Xu, H., Zang, C., Ripke, S., Bulik-Sullivan, B., Stahl, E., et al.; Schizophrenia Working Group of the Psychiatric Genomics Consortium; SWE-SCZ Consortium; Schizophrenia Working Group of the Psychiatric Genomics Consortium; SWE-SCZ Consortium (2014). Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. Am. J. Hum. Genet. *95*, 535–552.

Gutierrez-Arcelus, M., Lappalainen, T., Montgomery, S.B., Buil, A., Ongen, H., Yurovsky, A., Bryois, J., Giger, T., Romano, L., Planchon, A., et al. (2013). Passive and active DNA methylation and the interplay with genetic variation in gene regulation. eLife *2*, e00523.

Hindorff, 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.

Hirota, T., Takahashi, A., Kubo, M., Tsunoda, T., Tomita, K., Sakashita, M., Yamada, T., Fujieda, S., Tanaka, S., Doi, S., et al. (2012). Genome-wide association study identifies eight new susceptibility loci for atopic dermatitis in the Japanese population. Nat. Genet. *44*, 1222–1226.

Hulse, A.M., and Cai, J.J. (2013). Genetic variants contribute to gene expression variability in humans. Genetics *193*, 95–108.

Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A., et al.; International IBD Genetics Consortium (IIBDGC) (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature *491*, 119–124.

Kaplow, I.M., MacIsaac, J.L., Mah, S.M., McEwen, L.M., Kobor, M.S., and Fraser, H.B. (2015). A pooling-based approach to mapping genetic variants associated with DNA methylation. Genome Res. *25*, 907–917.

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.

Koren, A., Handsaker, R.E., Kamitaki, N., Karlić, R., Ghosh, S., Polak, P., Eggan, K., and McCarroll, S.A. (2014). Genetic variation in human DNA replication timing. Cell *159*, 1015–1026.

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al.; Roadmap Epigenomics Consortium (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

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.

Levo, M., and Segal, E. (2014). In pursuit of design principles of regulatory sequences. Nat. Rev. Genet. *15*, 453–468.

Levo, M., Zalckvar, E., Sharon, E., Dantas Machado, A.C., Kalma, Y., Lotam-Pompan, M., Weinberger, A., Yakhini, Z., Rohs, R., and Segal, E. (2015). Unraveling determinants of transcription factor binding outside the core binding site. Genome Res. *25*, 1018–1029.

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 (SAM) format and SAMtools. Bioinformatics *25*, 2078–2079.

Lin, X., Yang, H., Zhang, H., Zhou, L., and Guo, Z. (2013). A novel transcription mechanism activated by ethanol: induction of Slc7a11 gene expression via inhibition of the DNA-binding activity of transcriptional repressor octamer-binding transcription factor 1 (OCT-1). J. Biol. Chem. *288*, 14815–14823.

McCarroll, S.A., Huett, A., Kuballa, P., Chilewski, S.D., Landry, A., Goyette, P., Zody, M.C., Hall, J.L., Brant, S.R., Cho, J.H., et al. (2008). Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. Nat. Genet. *40*, 1107–1112.

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.

Parkes, M., Barrett, J.C., Prescott, N.J., Tremelling, M., Anderson, C.A., Fisher, S.A., Roberts, R.G., Nimmo, E.R., Cummings, F.R., Soars, D., et al.; Wellcome Trust Case Control Consortium (2007). Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. Nat. Genet. *39*, 830–832.

Rahmani, M., Péron, P., Weitzman, J., Bakiri, L., Lardeux, B., and Bernuau, D. (2001). Functional cooperation between JunD and NF-kappaB in rat hepatocytes. Oncogene *20*, 5132–5142.

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.

Sanborn, A.L., Rao, S.S., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl. Acad. Sci. USA *112*, E6456–E6465.

Sanduja, S., Blanco, F.F., and Dixon, D.A. (2011). The roles of TTP and BRF proteins in regulated mRNA decay. Wiley Interdiscip. Rev. RNA *2*, 42–57.

Schaub, M.A., Boyle, A.P., Kundaje, A., Batzoglou, S., and Snyder, M. (2012). Linking disease associations with regulatory information in the human genome. Genome Res. *22*, 1748–1759.

Schödel, J., Bardella, C., Sciesielski, L.K., Brown, J.M., Pugh, C.W., Buckle, V., Tomlinson, I.P., Ratcliffe, P.J., and Mole, D.R. (2012). Common genetic variants at the 11q13.3 renal cancer susceptibility locus influence binding of HIF to an enhancer of cyclin D1 expression. Nat. Genet. *44*, 420–425, S1–S2.

Sun, L.D., Xiao, F.L., Li, Y., Zhou, W.M., Tang, H.Y., Tang, X.F., Zhang, H., Schaarschmidt, H., Zuo, X.B., Foelster-Holst, R., et al. (2011). Genome-wide association study identifies two new susceptibility loci for atopic dermatitis in the Chinese Han population. Nat. Genet. *43*, 690–694.

Tewhey, R., Kotliar, D., Park, D.S., Liu, B., Winnicki, S., Reilly, S.K., Andersen, K.G., Mikkelsen, T.S., Lander, E.S., Schaffner, S.F., and Sabeti, P.C. (2016). Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. Cell. Published online June 2, 2016. http://dx.doi.org/10.1016/j.cell.2016.04.027.

van de Geijn, B., McVicker, G., Gilad, Y., and Pritchard, J.K.W.A.S.P. (2015). WASP: allele-specific software for robust molecular quantitative trait locus discovery. Nat. Methods *12*, 1061–1063, Advance Online.

Waszak, S.M., Delaneau, O., Gschwind, A.R., Kilpinen, H., Raghav, S.K., Witwicki, R.M., Orioli, A., Wiederkehr, M., Panousis, N.I., Yurovsky, A., et al. (2015). Population variation and genetic control of modular chromatin architecture in humans. Cell *162*, 1039–1050.

Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. Cell *158*, 1431–1443.

White, M.A., Myers, C.A., Corbo, J.C., and Cohen, B.A. (2013). Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. Proc. Natl. Acad. Sci. USA *110*, 11952–11957.

Yamada, Y., Nishida, T., Ichihara, S., Sawabe, M., Fuku, N., Nishigaki, Y., Aoyagi, Y., Tanaka, M., Fujiwara, Y., Yoshida, H., et al. (2011). Association of a polymorphism of BTN2A1 with myocardial infarction in East Asian populations. Atherosclerosis *215*, 145–152.