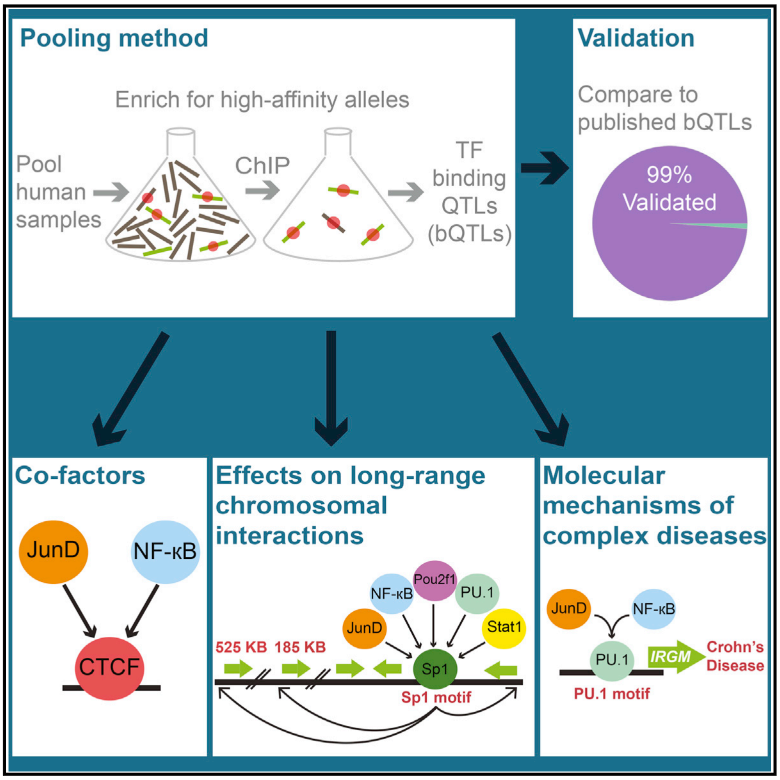
Resource

Cell

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

Graphical Abstract



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

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.

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

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