Systematic functional regulatory assessment of disease-associated variants

Konrad J. Karczewski[a,b,1], Joel T. Dudley[a,c,1,2], Kimberly R. Kukurba[b], Rong Chen[c], Atul J. Butte[c], Stephen B. Montgomery[b,d], and Michael Snyder[b,3]

[a]Biomedical Informatics Training Program, [b]Department of Genetics, [c]Division of Systems Medicine, Department of Pediatrics, and [d]Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305

Edited by Joseph R. Ecker, Salk Institute, La Jolla, CA, and approved April 24, 2013 (received for review November 3, 2012)

Genome-wide association studies have discovered many genetic loci associated with disease traits, but the functional molecular basis of these associations is often unresolved. Genome-wide regulatory and gene expression profiles measured across individuals and diseases reflect downstream effects of genetic variation and may allow for functional assessment of disease-associated loci. Here, we present a unique approach for systematic integration of genetic disease associations, transcription factor binding among individuals, and gene expression data to assess the functional consequences of variants associated with hundreds of human diseases. In an analysis of genome-wide binding profiles of NFκB, we find that disease-associated SNPs are enriched in NFκB binding regions overall, and specifically for inflammatory-mediated diseases, such as asthma, rheumatoid arthritis, and coronary artery disease. Using genome-wide variation in transcription factor-binding data, we find that NFκB binding is often correlated with disease-associated variants in a genotype-specific and allele-specific manner. Furthermore, we show that this binding variation is often related to expression of nearby genes, which are also found to have altered expression in independent profiling of the variant-associated disease condition. Thus, using this integrative approach, we provide a unique means to assign putative function to many disease-associated SNPs.

systems biology | regulatory genomics | translational bioinformatics

Elucidation of functional mechanisms underlying genetic associations with phenotypic traits is a fundamental problem in biology and its translation to medicine. Genome-wide association studies (GWAS) have identified many genetic variants associated with diseases (1), but such approaches rely on "tag" single nucleotide polymorphisms (SNPs) found on DNA microarrays. Whereas these SNPs may lie within or near genes or other functional regions, their specific functional relationships to the biology of disease are not necessarily determined through genetic association alone (2).

Integrative genomics can provide an approach to bridge the gap between genotype and phenotype. Regulatory features across hundreds of transcription factors (TFs) and dozens of cell lines have been mapped extensively using ChIP-Seq (Chromatin Immunoprecipitation followed by high-throughput sequencing) by the ENCODE project (3). We expect that polymorphisms that affect transcription factor binding can have a tremendous influence on disease (4), as the differences in TF binding that lead to downstream differences in expression may be the underlying cause of the disease association of the SNPs. Molecular profiling data measuring DNA, TF binding, and mRNA expression variation across individuals are recently published (5), and large compendia of mRNA expression profiles of disease states are available in the public domain (6). Integrative analysis of these functional biology-rich sources of data may suggest putative function for previously unannotated disease-associated SNPs.

Previous approaches to the study of regulatory variation have focused on single diseases and regions or have taken a genome-wide approach, but have not systematically explored allele-specific effects of DNA binding. For instance, studies have focused on regions such as 9p21, a well-studied gene desert associated with Coronary Artery Disease (CAD), which has been shown to harbor many enhancers and disease risk alleles in this region that disrupt TF binding sites (7). Adrianto et al. used an unbiased approach, performing a GWAS with a functional follow-up experiment to show that a risk-conferring variant affects NFκB binding (8). Genome-wide methods have successfully related diseases to transcription factors, based on genome-wide binding data and disease association data from GWAS (4, 9). However, these approaches have not systematically explored allelic effects of disease-associated variants. A recent analysis across cell types has shown allele-specific DNase hypersensitivity within each cell type (10), but does not explore the direct effect of natural human variation across individuals on regulatory features. A similar analysis was performed across chromatin marks, showing how variants associated with breast cancer can affect chromatin affinity and gene expression (11). These studies have provided evidence that variants in TF binding sites have roles in disease specific to the biology of the TF.

In this work, we focus on the functional role of variants in transcription factor binding sites in human disease. As a case study, we explore variants in the binding regions of NFκB, which is a crucial regulator of inflammation and has been implicated in many diseases, including autoimmune diseases and cancer (12). Additionally, variation in NFκB binding has been extensively mapped and correlated with variants in motifs and the binding site (5, 13). Here, we use these natural variation data to investigate properties of disease-associated SNPs in NFκB binding regions and provide putative functional explanations for their disease mechanisms, through genotype-specific and allele-specific binding events.

Results

NFκB Binding Regions Are Enriched for Disease-Associated SNPs. We mapped a compendium of 66,128 common (dbSNP 135, ≥1% overall minor allele frequency, MAF) disease-associated SNPs (14, 15) to a set of 15,522 NFκB binding regions found in lymphoblastoid cell lines from 10 individuals (5) (Fig. 1, *Top*). These binding regions span 15.1 Mb and contain 60,595 common SNPs, which allowed for a reduced set of candidate functional variants.

We found 797 established disease-associated SNPs (representing 144 diseases) in regions bound by NFκB, a significant overrepresentation (2.25-fold; Fig. 2*A*) compared with all common variants (Fisher's exact $P = 4.2 \times 10^{-90}$). This enrichment is even more pronounced for stringent disease associations, including genome-wide significant variants (GWAS P value $<10^{-7}$) and variants that have been replicated in multiple studies and multiple ethnicities. These enrichments and trends are similar to

Author contributions: K.J.K., J.T.D., and M.S. designed research; K.J.K., J.T.D., and K.R.K. performed research; K.J.K., R.C., A.J.B., and S.B.M. contributed new reagents/analytic tools; K.J.K., J.T.D., and K.R.K. analyzed data; and K.J.K., J.T.D., and M.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]K.J.K. and J.T.D. contributed equally to this work.

[2]Present address: Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

[3]To whom correspondence should be addressed. E-mail: mpsnyder@stanford.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1219099110/-/DCSupplemental.

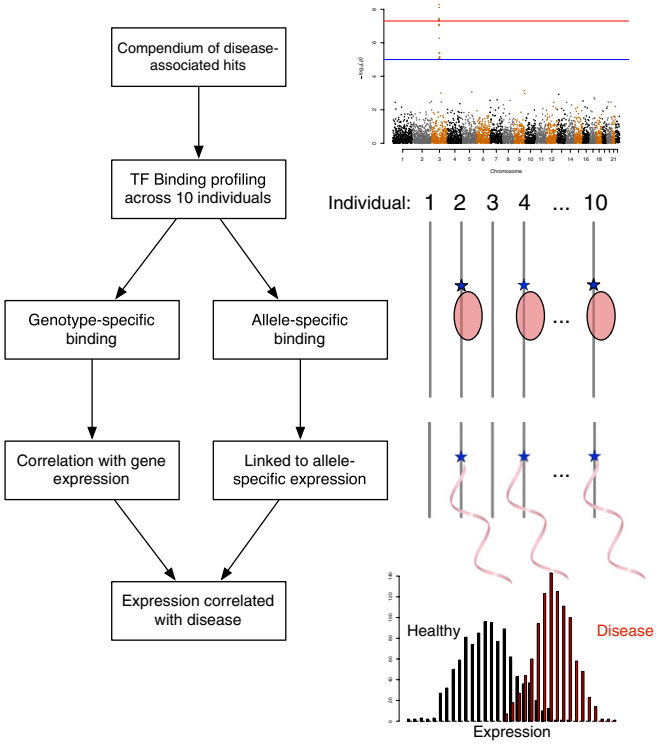

Fig. 1. Regulatory variant assessment pipeline. We intersected a compendium of disease-associated variants with NFκB binding sites. We assessed the effect of these and variants in LD with these variants on binding and expression and linked these expression effects back to the same disease.

those for general regulatory elements, such as variants in sites bound by RNA polymerase II (PolII) (Fig. S1). These associations are not biased by genomic parameters, such as minor allele frequency or distance to transcription start site (Fig. S2). Additionally, this enrichment is even stronger for variants in linkage disequilibrium with variants in NFκB binding sites (Fig. 2*B*).

Next, we evaluated the presence of genetic associations in NFκB regions on a per-disease basis. We found that SNPs associated with primarily inflammatory and autoimmune diseases and cancers, including rheumatoid arthritis, asthma, and lymphoma, were highly overrepresented [false discovery rate (FDR) < 0.1] in NFκB binding regions (Fig. 2*C* and Dataset S1) compared with a background of the overall compendium of disease-associated SNPs. Additionally, SNPs that reach genome-wide significance (GWAS *P* value $<10^{-7}$) associated with systemic lupus erythematosus, primary biliary cirrhosis, and rheumatoid arthritis are also enriched in NFκB regions (Fig. 2*C*). These disease signals are TF specific and not a general property of regulatory features, as demonstrated by the enrichment of different diseases for PolII binding regions (Fig. S1).

Finally, to determine if disease-associated variants are associated with the most occupied sites, we examined whether the NFκB binding regions that harbor disease-associated SNPs have higher ChIP-Seq signals compared with the background of all NFκB binding regions. We found that the disease-associated signals are 29.3% higher on average than all NFκB binding sites (Fig. 2*D*), and this increase is not due to differences in GC content (Fig. S3). Additionally, we find that the ChIP-Seq strength is even higher for binding sites with high-confidence disease-associated variants. Thus, disease mechanisms may be identified among the stronger, more highly occupied NFκB binding sites.

Disease-Associated SNPs in and Linked to NFκB Regions Are Highly Functional. We explored the association properties of disease-associated variants in NFκB binding regions. On average, we find that the effect size from the originating GWAS [as measured by OR, odds ratio (OR) or likelihood ratio (LR)] is higher for variants in NFκB binding regions than the average disease-associated SNP (Fig. 3*A*; mean log OR: 0.508 vs. 0.448; *t* test $P = 1.7E\text{-}4$). These effects are specifically pronounced for variants associated with gliomas, rheumatoid arthritis, and systemic lupus erythematosus (Table 1). Additionally, SNPs in NFκB regions in linkage disequilibrium (LD) with disease-associated variants are enriched for inclusion in NFκB motifs (over all SNPs in NFκB regions; 1.54-fold, $P = 1.3e\text{-}06$), which is even higher for stringent (disease association $P < 1e\text{-}7$) associated SNPs (3.17-fold; $P = 3.4e\text{-}13$).

We also observe that disease-associated SNPs in NFκB binding regions are more pleiotropic (i.e., generally associated with more diseases) than the collection of known disease-associated SNPs (Fig. 3*B* and Fig. S4; 1.36 vs. 1.12; Mann-Whitney *U*, *P* value = 2.7e-7). This observation likely reflects the broad role of NFκB in many inflammatory-related, but subtly different biological processes.

SNPs in NFκB Binding Regions Suggest a Mechanism for the Biology of Disease. To systematically functionally annotate to disease-associated SNPs, we developed a pipeline to discover putative SNPs that may be associated with an effect on NFκB binding (Fig. 1). We identified candidate functional disease GWAS SNPs using genotype and NFκB binding information from eight individuals (5), where SNPs associating with differential NFκB binding in an genotype-specific (cross-individual) or allele-specific (heterozygous) fashion are presumed to be functional. In an assessment of 5,497 SNPs in NFκB binding regions in linkage disequilibrium [$r^2 > 0.3$ in Hapmap European (CEU) samples] with one or more disease-associated SNPs, we found a total of 317 SNPs whose genotype was associated with differential NFκB binding across individuals (nominal $P < 0.05$ and overall FDR < 0.1; Dataset S2). We find five additional variants that show allele-specific NFκB binding, within a heterozygous site in a single individual, (binomial $P < 0.05$ in at least one sample) in LD with disease-associated SNPs.

As an example, rs171407, a SNP previously reported to be associated with breast cancer (16) and linked ($r^2 = 0.812$) to rs35683, a variant associated with type 2 diabetes (17), shows significant genotype-specific association with NFκB binding in the eight individuals queried ($r = -0.88$; $P = 0.0038$; Fig. 4*A*). Additionally, we find an allele-specific binding event at rs12588969 across four individuals (31.1% reference allele; pooled binomial $P = 1.28 \times 10^{-4}$; Fig. 4*B*), which may provide a functional explanation for two disease-associated variants: rs10137035 ($r^2 = 0.464$), associated with systemic lupus erythematosus (18), and rs941726 ($r^2 = 0.36$), associated with diffuse large b-cell lymphoma (19).

Overall, we find that variants in NFκB binding regions that are linked to disease-associated variants are 43.6% more likely to be B-SNPs, or SNPs in binding regions that are significantly associated with TF binding (nominal $P < 0.01$), than variants in binding sites not in LD with GWAS SNPs (Fisher's exact $P = 0.0019$; Fig. 4*C*). However, only 20–25% of the variants that affect binding are linked to disease-associated variants (Fig. S5), reflecting the potential for additional mechanisms yet undiscovered.

Regulatory SNPs Are Linked to Disease-Associated Molecular Phenotypes. Next, we determined if disease-associated variants implicated in variable NFκB binding could be linked to downstream functional regulatory effects on the expression of nearby genes (Fig. 1, *Middle*). Considering all genes within 100 kb to be potential targets, of the 317 variants associated with NFκB binding, we found 64 where the binding strength was also correlated with expression of nearby genes (nominal $P < 0.05$ and overall FDR < 0.1). For instance, rs3784275 is linked ($r^2 = 0.866$) to rs12702 (a variant associated with diabetic nephropathy) (20): rs3784275 is correlated with NFκB binding ($r = 0.889$; $P = 0.0032$), which in turn is correlated with expression of the uncharacterized protein KIAA1737 ($r = 0.802$; $P = 0.01$; Fig. S6*A*).

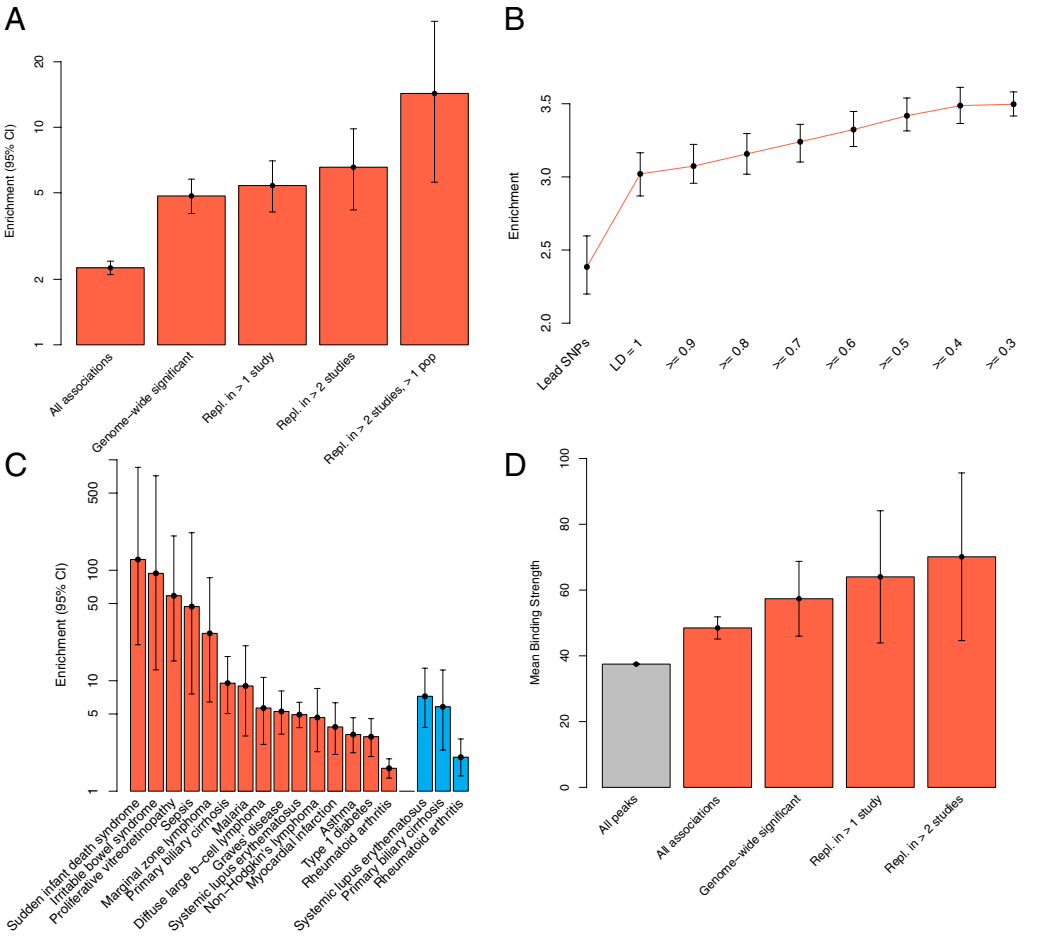

Fig. 2. NFκB regions are enriched for disease-associated variants. (*A*) Fold-enrichments for disease-associated variants in NFκB regions: all associations and genome-wide significant variants include variants with nominal $P < 0.01$ and $P < 10^{-7}$, respectively, in the original study. Error bars reflect the 95% confidence interval of Fisher's exact test. (*B*) This enrichment is more pronounced when considering variants in LD with disease-associated variants. (*C*) Initial (lead SNP) enrichment analysis was repeated on a per-disease basis. Enrichment is shown for all associations relative to all diseases (red) and genome-wide significant associated variants (blue). (*D*) NFκB binding sites with disease-associated variants are stronger binding peaks than the average NFκB binding site. Error bars are shown based on *t* test 95% confidence interval.

SYSTEMS BIOLOGY

To further explore the possibility that the downstream functional regulatory effects of SNPs mediated by NFκB binding variability contribute directly to the pathophysiology of disease, we evaluated the expression patterns of NFκB binding genes across 143 human disease conditions (Fig. 5). Many diseases exhibit broad patterns of dysregulation of NFκB binding genes. Among these are diseases for which NFκB is known to play a substantial role in the molecular pathophysiology of the disease, such as asthma, Alzheimer's, and glioblastoma. However, other diseases for which NFκB is not thought to play a substantial role, such as cardiomyopathy and polycystic ovary syndrome, also exhibit similar patterns of broad dysregulation of NFκB binding

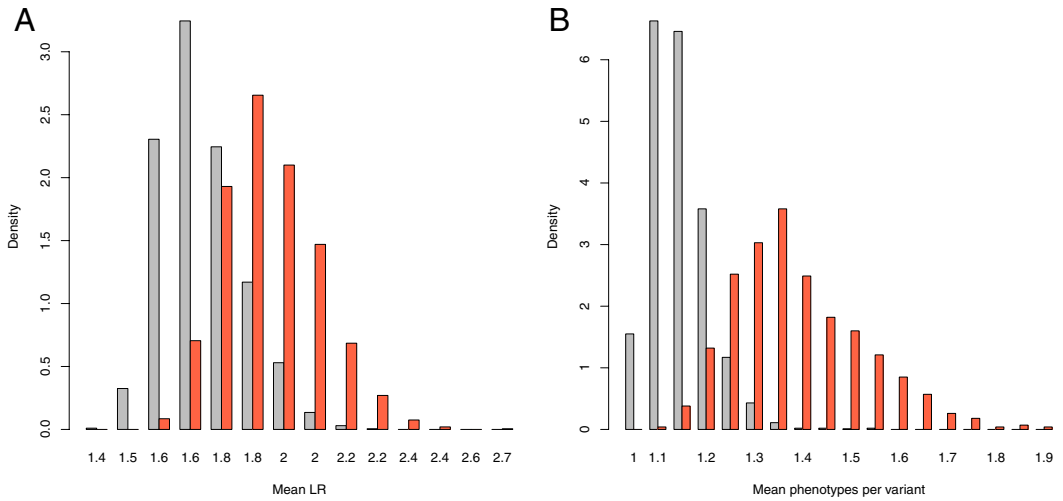

Fig. 3. Disease-associated variants in NFκB regions are highly functional. (*A*) Disease-associated variants in NFκB regions (red) have higher odds ratios than the average disease-associated variants (gray). Means of 1,000 sampling distributions are plotted. (*B*) Disease-associated variants in NFκB regions (red) are associated with more diseases per variant (are more pleiotropic) than the collection of disease-associated variants (gray).

Table 1. Variants in NFκB regions have strong effects in particular diseases

Broad phenotype	NFκB mean LR	Mean LR	*t* test *P*	Wilcox *P*
Glioma	0.9692	0.6472	0.1031	0.0419
Rheumatoid arthritis	0.3679	0.3054	0.0118	0.0013
Systemic lupus erythematosus	0.4276	0.3587	0.0410	0.0010

For these diseases, the mean likelihood ratio (LR) for variants in NFκB regions is higher than the average LR for variants associated with the disease.

genes. However, we note that a suspected role for NFκB in the pathophysiology of cardiomyopathy is recently emerging in the literature (21).

In several cases, NFκB binding genes dysregulated in a disease can already be linked to genetic variants associated with the same disease falling within associated regions of correlated NFκB binding activity (Fig. 1, *Bottom*). For example, IL-12B is found to be dysregulated in inflammatory bowel disease (IBD) (22–24), and a SNP in IL-12B rs6871626 was associated with IBD phenotype in an independent GWAS (25). Allelic variation in rs12651787 (linked to rs6871626: $r^2 = 0.72$) is significantly associated with NFκB binding variability in a NFκB binding region upstream of IL-12B ($r = -0.827$; $P = 0.011$), and binding variability in this region is significantly associated with RNA transcript levels of IL-12B ($r = 0.78$; $P = 0.021$; Fig. S6*B*). Therefore, genetic variation in regions linked to rs12651787 may serve as etiological factors for IBD through downstream effects on the regulation of IL-12B expression.

Discussion

In this study, we explore the physiological effects of regulatory variants in NFκB binding sites using GWAS information. Our work is significantly different from that recently reported through ENCODE studies that primarily mapped binding and open chromatin information with disease-associated GWAS hits. By using genotype information from multiple individuals, we obtain functional information about the effects of allelic variation on NFκB binding, which in turn is correlated with disease and expression information. In this manner, a much stronger association can be made between genetic variation and biological function.

By correlating NFκB binding regions with disease associations, we propose potential genetic mechanisms for the etiology of many inflammatory and immune-related diseases. We confirmed known diseases (such as rheumatoid arthritis, asthma, and lymphomas) associated with NFκB binding and suggest additional associations (sudden infant death syndrome; SIDS). In particular, this result lends support to the link between inflammation and SIDS (26) and suggests that genetic variation in NFκB regions may contribute to its pathophysiology.

Increased pleiotropy at disease-associated variants highlights the complexity of regulatory variation (Fig. 3*B*). These effects are typically more subtle than rare coding variants, such as those found in Mendelian disorders that have large individual, but but highly specific, effects. Instead, the variants uncovered in this study may act through perturbation of a network involved in many biological processes (27), thereby leading to many possible phenotypes. The distinct phenotypes may depend on any number of environmental triggers or complex epistatic genetic interactions.

This study describes in detail the role of variants in NFκB binding sites in potential disease mechanisms. The mapping of additional variation datasets for other transcription factors by the ENCODE

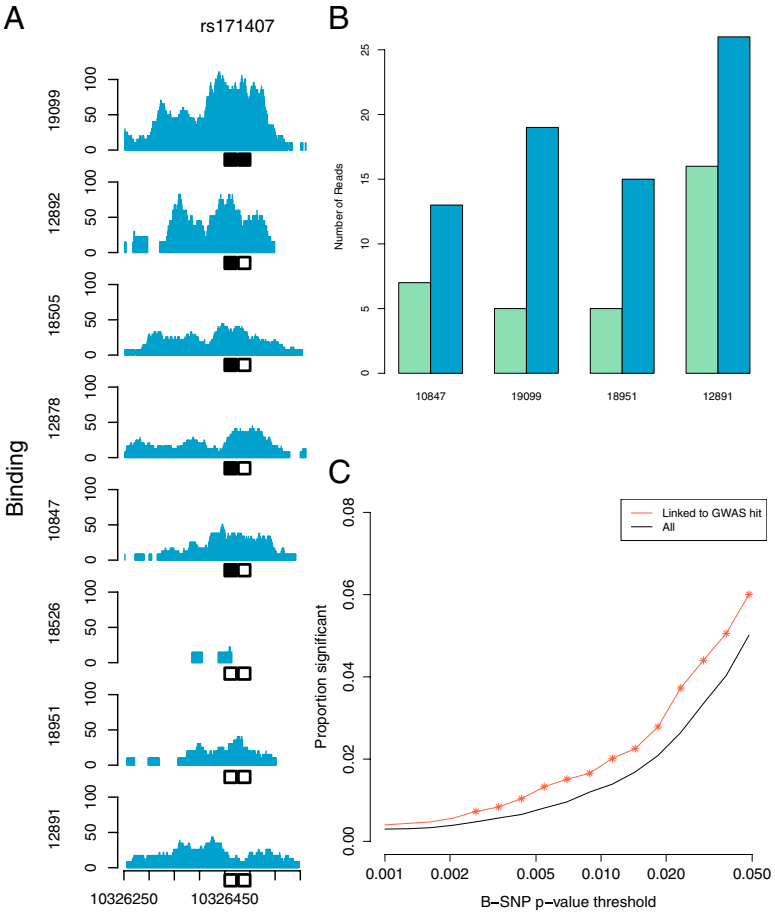

Fig. 4. Disease-associated variants contribute to regulatory effects. (*A*) Regulatory variant effect associated with disease: rs171407, a variant associated with breast cancer and linked to a variant (rs35683) associated with type 2 diabetes, is associated with an NFκB binding site in which it resides. Eight individuals are shown and their variant state is shown by the boxes below. (*B*) Variant (rs12588969) in an NFκB binding region that shows allele-specific binding of the nonreference allele (blue) over the reference allele (green) in four individuals. This variant is linked to rs10137035 and rs941726, which are associated with systemic lupus erythematosus and diffuse large b-cell lymphoma, respectively. (*C*) Variants in NFκB regions that are linked to disease-associated variants are more likely to be B-SNPs, or associated with NFκB binding at a number of *P* value thresholds, than those not in LD with disease-associated hits (*$P < 0.05$ denotes a significant increase between linked and unlinked variants).

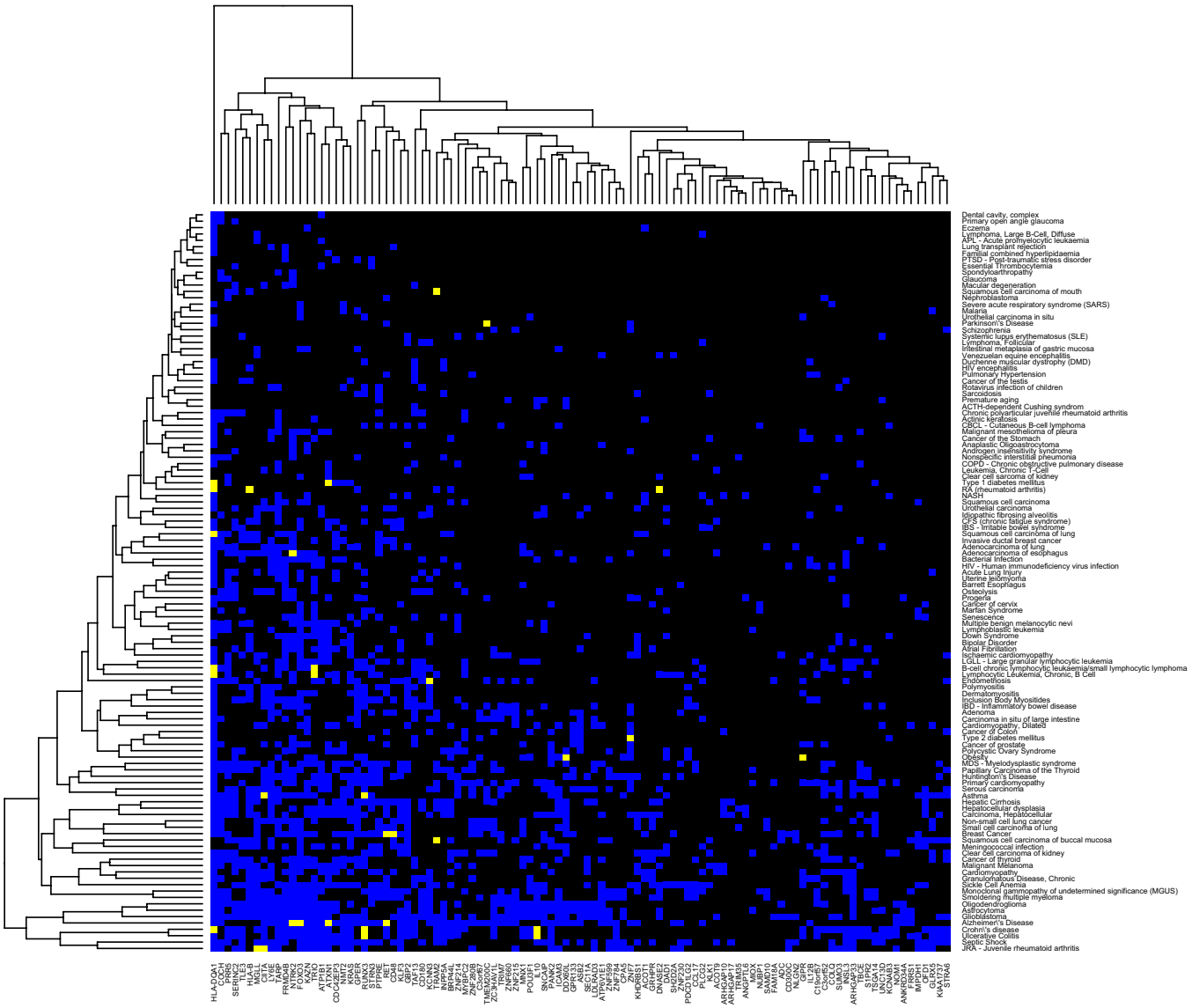

Fig. 5. Expression landscape of NFκB binding genes across human diseases. Clustered heatmap shows the binary differential expression profile of 108 NFκB binding genes (columns) across 143 distinct human disease conditions (rows). Blue-colored cells indicate that the gene is observed to be differentially expressed in the disease condition relative to normal controls, and black indicates a lack of observed differential expression. Yellow-colored cells indicate that the gene is observed to be differentially expressed in the disease and that variants near that gene associated with NFκB binding are linked to disease-associated hits for the same disease. Many diseases having a substantial and well-established role for NFκB in their pathologies, such as asthma and glioblastoma, exhibit similar NFκB binding gene expression profiles as diseases not traditionally thought to have a substantial basis in NFκB regulatory activity, such as cardiomyopathy and polycystic ovary disease.

project and others will enable a comprehensive assessment of regulatory information in disease. Additionally, as the regulatory code is deciphered, the role of cooperativity among factors in disease will need to be elucidated.

It is expected that a large number of individual genomes will be sequenced in the near future, and we expect that the need to evaluate and interpret association of binding SNPs with noncoding variants will continue to gain in importance, especially given most disease-associated variants lie outside of gene coding regions (1). Therefore, it is likely that comprehensive, integrative analysis of regulatory variants will be crucial for fully characterizing the etiology of complex diseases. DNA variation in NFκB binding regions associated with patterns of dysregulated NFκB binding genes may highlight regions likely to harbor novel genetic risk variants underlying susceptibility to various diseases, and may serve as a means for prioritizing regions for genetic analysis. These molecular connections may also provide insight into possible biological mechanisms by which genetic risk variants identified through association studies are connected to functional molecular phenotypes of disease. Thus, variants that lie in binding regions that vary among individuals (i.e., B-SNPs) and their association with disease loci as described in this study is expected to be extremely valuable for clinical genomics and the annotation of personal genomes.

Methods

Data Sources. Data on 121,543 disease-SNP associations ($P < 0.01$) from the Varimed database were used as in refs. 14 and 15, which, where possible, include odds ratios and likelihood ratios from initial publications and 1,410 phenotypes mapped to MeSH terms. When a protective odds ratio or likelihood ratio is reported in the literature, it is transformed to a risk allele using its inverse (risk OR = 1/protective). These variants were filtered to

variants in dbSNP 135 with ≥1% overall MAF resulting in 66,128 disease-associated SNPs. ChIP-Seq data for NFκB and PolII for lymphoblast cell lines derived from 10 individuals (for 8 of which individual genome sequences were available), including quantitative binding information (normalized for sample coverage) across individuals were obtained from ref. 5. All analyses were performed using dbSNP release 135 and hg19 coordinates. Variant annotations were obtained from dbSNP135 annotations from University of California Santa Cruz. All statistical analysis methods were performed using R statistical software (2.15.1).

TF-Disease Enrichments. Variants from the disease-association database were mapped onto NFκB binding regions lifted over from hg18 to hg19, and this intersection retained 144 diseases (Dataset S3). Enrichments for disease-associated SNPs in binding regions and other functional classes were ascertained by Fisher's exact tests. Correction for multiple hypothesis testing was assessed using *q*-value FDR analysis (the R package, qvalue) (28). Additionally, per-disease enrichments were corrected using permutation analysis: SNPs were dissociated from their associated diseases and significant enrichments were required to have $q < 0.1$ and $P_{perm} < 0.1$.

Simulated backgrounds were generated from 1,000 random samples of variants in dbSNP135, limited to variants of at least 1% MAF, whose distribution was matched to the joint distribution of the MAF and distance to transcription start site (TSS) of the disease-associated variant database. The number of disease-associated variants in TF binding sites was then compared with this distribution to estimate an empirical *P* value. Simulations were also run in reverse: the number of disease-associated variants in TF binding sites was compared with the variants sampled from a matched distribution of MAF and distance to TSS of variants in TF binding sites.

SNP, Binding, and Expression Association. Associations between individual SNPs and binding strengths were tested using Pearson correlation of number of nonreference alleles to the quantitative measure of NFκB binding. Associations between binding and expression were ascertained by Pearson correlation between NFκB binding and expression (reads per kilobase per million) obtained from ref. 5. Permutation testing was run for the joint distribution between the effect of variants on binding and binding on expression: for each variant–gene combination, 1,000 permutations of the variant, binding, and expression were generated and tested as above. We generated a distance metric based on the sum of correlation values ($r_{binding}^2 + r_{expression}^2$) and compared the true value against this distribution to estimate an empirical *P* value (Fig. S7).

Allele-Specific Binding. ChIP-Seq reads for all 10 individuals were remapped to hg19 using BWA (0.6.1) and filtered for PCR duplicates using Picard (1.72). Variant calls were obtained from the 1000 Genomes Project and converted to hg19 coordinates using liftOver. Allele-specific binding (ASB) was determined on a per-heterozygote per-individual basis for the 10 individuals, as in ref. 29. Reads were filtered to be above MAQ 30 mapping quality. For each individual, a binomial probability of success was determined based on the probability that a reference allele maps to the genome compared with a nonreference allele. Significant ASB events were determined as sites with at least five reads per allele, binomial $P < 0.01$, and where the nonreference allele was overrepresented (to minimize reference sequence bias). Allele-specific expression (ASE) was similarly determined using reads from the transcriptome (RNA-Seq) of each individual.

Disease–Gene Association. We obtained disease vs. normal gene expression profiles for 300 human diseases using methods described previously (6). In brief, we identified published microarray studies in the National Center for Biotechnology Information Gene Expression Omnibus (GEO) and European Bioinformatics Institute ArrayExpress databases relevant to human disease. Each study is annotated with controlled disease and tissue terms selected from the Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT) and National Cancer Institute thesaurus vocabularies, respectively. Only those experiments having normal, tissue-matched controls measured in the same experiment were retained. For each disease vs. control experiment, we estimate the set of differentially expressed genes using RankProd software with a 5% false discovery rate threshold. One hundred sixteen SNOMED-CT terms represented in the expression profiles using the Unified Medical Language System (UMLS) were mapped to disease Medical Subject Heading (MeSH) terms annotations in Varimed, where the overlap thereof was 89 diseases (Dataset S3).

To evaluate the NFκB binding gene profiles among diseases, we represented each disease as binary expression vector of length *n*, where $n = 108$ representing the 108 NFκB binding genes. The *i*th position in the vectors represents a distinct binding gene g_i, and if a disease d_i is found to differentially express g_i, then $g_i = 1$, otherwise $g_i = 0$. We performed the heatmap cluster analysis by first estimating the Manhattan distance matrix between each disease vector pair, followed by agglomerative hierarchical clustering using the average linkage method. We further annotated the heatmap by identifying genes differentially expressed in a disease that are also linked to NFκB binding regions harboring disease susceptibility variants identified for the same disease through genetic association studies.

ACKNOWLEDGMENTS. K.J.K. is supported by the National Science Foundation Graduate Research Fellowship Program. K.J.K. and J.T.D. are supported by National Institutes of Health (NIH) Training Grant LM007033. J.T.D., R.C., and A.J.B. are supported by the Hewlett Packard Foundation and Lucile Packard Foundation for Children's Health. K.R.K. is supported by a National Defense Science and Engineering Graduate Fellowship. M.S. is supported by grants from the NIH.

1. Hindorff LA, et al. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* 106:9362–9367.
2. Green ED, Guyer MS; National Human Genome Research Institute (2011) Charting a course for genomic medicine from base pairs to bedside. *Nature* 470:204–213.
3. ENCODE Project Consortium et al. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489:57–74.
4. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M (2012) Linking disease associations with regulatory information in the human genome. *Genome Res* 22: 1748–1759.
5. Kasowski M, et al. (2010) Variation in transcription factor binding among humans. *Science* 328:232–235.
6. Dudley JT, Tibshirani R, Deshpande T, Butte AJ (2009) Disease signatures are robust across tissues and experiments. *Mol Syst Biol* 5:307.
7. Harismendy O, et al. (2011) 9p21 DNA variants associated with coronary artery disease impair interferon-γ signalling response. *Nature* 470:264–268.
8. Adrianto I, et al. (2011) Association of a functional variant downstream of TNFAIP3 with systemic lupus erythematosus. *Nat Genet* 43:253–258.
9. Ernst J, et al. (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473:43–49.
10. Maurano MT, et al. (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337:1190–1195.
11. Cowper-Sal Lari R, et al. (2012) Breast cancer risk–associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat Genet* 44:1191–1198.
12. Tak PP, Firestein GS (2001) NF-κB: A key role in inflammatory diseases. *J Clin Invest* 107:7–11.
13. Karczewski KJ, et al. (2011) Cooperative transcription factor associations using regulatory variation information. *Proc Natl Acad Sci USA* 108:13353–13358.
14. Ashley EA, et al. (2010) Clinical assessment incorporating a personal genome. *Lancet* 375:1525–1535.
15. Chen R, Davydov EV, Sirota M, Butte AJ (2010) Non-synonymous and synonymous coding SNPs show similar likelihood and effect size of human disease association. *PLoS ONE* 5:e13574.
16. Dossus L, et al. (2008) Polymorphisms of genes coding for ghrelin and its receptor in relation to anthropometry, circulating levels of IGF-I and IGFBP-3, and breast cancer risk: A case-control study nested within the European Prospective Investigation into Cancer and Nutrition (EPIC). *Carcinogenesis* 29:1360–1366.
17. Garcia EA, et al. (2009) The role of ghrelin and ghrelin-receptor gene variants and promoter activity in type 2 diabetes. *Eur J Endocrinol* 161:307–315.
18. Sandling JK, et al. (2011) A candidate gene study of the type I interferon pathway implicates IKBKE and IL8 as risk loci for SLE. *Eur J Hum Genet* 19:479–484.
19. Wang SS, et al. (2011) Variations in chromosomes 9 and 6p21.3 with risk of non-Hodgkin lymphoma. *Cancer Epidemiol Biomarkers Prev* 20:42–49.
20. Savage DA, et al. (2008) Genetic association analyses of non-synonymous single nucleotide polymorphisms in diabetic nephropathy. *Diabetologia* 51:1998–2002.
21. Lorenzo O, et al. (2011) Potential role of nuclear factor κB in diabetic cardiomyopathy. *Mediators Inflamm* 2011:652097.
22. Glas J, et al. (2012) Analysis of IL12B gene variants in inflammatory bowel disease. *PLoS ONE* 7:e34349.
23. Bouma G, Strober W (2003) The immunological and genetic basis of inflammatory bowel disease. *Nat Rev Immunol* 3:521–533.
24. Parrello T et al. (2000) Up-regulation of the IL-12 receptor β2 chain in Crohn's disease. *J Immunol* 165:7234–7239.
25. Anderson CA, et al. (2011) Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet* 43:246–252.
26. Blood-Siegfried J (2009) The role of infection and inflammation in sudden infant death syndrome. *Immunopharmacol Immunotoxicol* 31:516–523.
27. Schadt EE, Björkegren JLM (2012) NEW: Network-enabled wisdom in biology, medicine, and health care. *Sci Transl Med* 4:115rv1.
28. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 100:9440–9445.
29. Montgomery SB, et al. (2010) Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* 464:773–777.