Cooperative transcription factor associations discovered using regulatory variation

Konrad J. Karczewski[a,b], Nicholas P. Tatonetti[a,b], Stephen G. Landt[b], Xinqiong Yang[b], Teri Slifer[b], Russ B. Altman[b], and Michael Snyder[b,1]

[a]Biomedical Informatics Training Program and [b]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305

Regulation of gene expression at the transcriptional level is achieved by complex interactions of transcription factors operating at their target genes. Dissecting the specific combination of factors that bind each target is a significant challenge. Here, we describe in detail the Allele Binding Cooperativity test, which uses variation in transcription factor binding among individuals to discover combinations of factors and their targets. We developed the ALPHABIT (a large-scale process to hunt for allele binding interacting transcription factors) pipeline, which includes statistical analysis of binding sites followed by experimental validation, and demonstrate that this method predicts transcription factors that associate with NFκB. Our method successfully identifies factors that have been known to work with NFκB (E2A, STAT1, IRF2), but whose global coassociation and sites of cooperative action were not known. In addition, we identify a unique coassociation (EBF1) that had not been reported previously. We present a general approach for discovering combinatorial models of regulation and advance our understanding of the genetic basis of variation in transcription factor binding.

functional genomics | systems biology

O ver the last several decades, it has become increasingly clear that the control of gene expression is due to the complex interactions of different transcription factors (TFs) working together to regulate RNA polymerase activity at promoters (1, 2). Although one factor may be a global regulator of a single process, it may function with other TFs to achieve precise regulation at different loci. Identifying the factors that work together to regulate each gene is a fundamental problem in biology.

A variety of approaches have been used to measure TF coassociation. In vitro and in vivo biochemical assays have been used to detect protein–protein interactions, but these are plagued by trade-offs between sensitivity and specificity (3, 4). Moreover, in vitro assays do not always reflect the events that occur in vivo. Computational approaches have been successful in predicting binding sites by analyzing motifs in promoter regions in the context of expression data, but these are noisy and do not consider condition-specific binding events (5). Coassociation of TF binding sites can also be used to predict factors that work together (6, 7), but such approaches lack functional information about specific interactions and typically require large numbers of coassociated sites.

We have recently suggested a unique approach, the Allele Binding Cooperativity (ABC) test, which uses binding variation among individuals to identify TF coassociation (8, 9). We hypothesized that variation in TF binding can occur because of sequence variation in associated TF binding sites and motifs. By searching for motifs in the binding regions for a factor of interest, the covariance of the associated motif can be correlated with binding of the factor across individuals. We demonstrated this phenomenon in a preliminary proof-of-concept, but did not rigorously examine many important parameters of the ABC test, nor did we verify that the associated motifs were bound by the predicted factor.

Here, we demonstrate that the ABC test can be used to systematically identify coassociated TFs and the targets they regulate. We examine the parameters for optimizing this approach and

describe the ALPHABIT (A Large scale Process to Hunt for Allele Binding Interacting Transcription factors) computational and experimental pipeline to identify targets from variation data. We apply this method to the identification of novel binding partners of NFκB (p65), whose binding sites were mapped in 10 individuals using ChIP-Seq. (8). NFκB is a key regulator of many cellular processes including inflammation, immune response, and cellular proliferation; ~7.5% of NFκB binding sites were found to be variable between any two individuals, but only 0.7% of these variable sites contained SNPs within the canonical NFκB binding motif (binding SNPs or B-SNPs). Another 31.5% of variable binding sites contain a SNP within the NFκB peak, but not in the NFκB motif itself; it is plausible that many of these SNPs lie in the binding sites of TFs that coassociate with NFκB at those targets. We used the ABC test and ALPHABIT pipeline to identify TFs (EBF, STAT1, E2A, and IRF2) that may bind cooperatively with NFκB and identified loci where they coassociate.

Results

Overview of the ABC Approach and the ALPHABIT Pipeline. NFκB binding varies at many sites in cell lines from different individuals, but few of these lines exhibit variability in the NFκB motif at the variable regions. We therefore reasoned that SNPs in the motifs of cooperatively binding factors may affect binding of NFκB (Fig. 1A) and established the ALPHABIT pipeline to investigate binding partners of NFκB (Fig. 1B). In this approach, we first search for binding motifs of all 146 TFs from the JASPAR database (10) in variable NFκB binding regions (8). At each of these regions, we find the difference of a motif of interest between each pair of individuals [weighted using a position weight matrix (PWM) metric; see *Methods*], as well as their fold-difference of NFκB binding. We then aggregate these data across binding sites and correlate the motif differences with binding differences. The interacting factors with the most significant correlations are candidates to be NFκB partners and are then tested for coassociation by ChIP-Seq (see below).

Identification of Motifs Whose Variance Correlates with NFκB Binding.

We used the ALPHABIT pipeline to analyze variable NFκB binding regions for motifs of possible coassociated TFs. Using the JASPAR database, we searched for motifs that contained SNPs whose match to the PWM correlated with NFκB binding. We term such SNPs "cB-SNPs," for cooperative B-SNPs
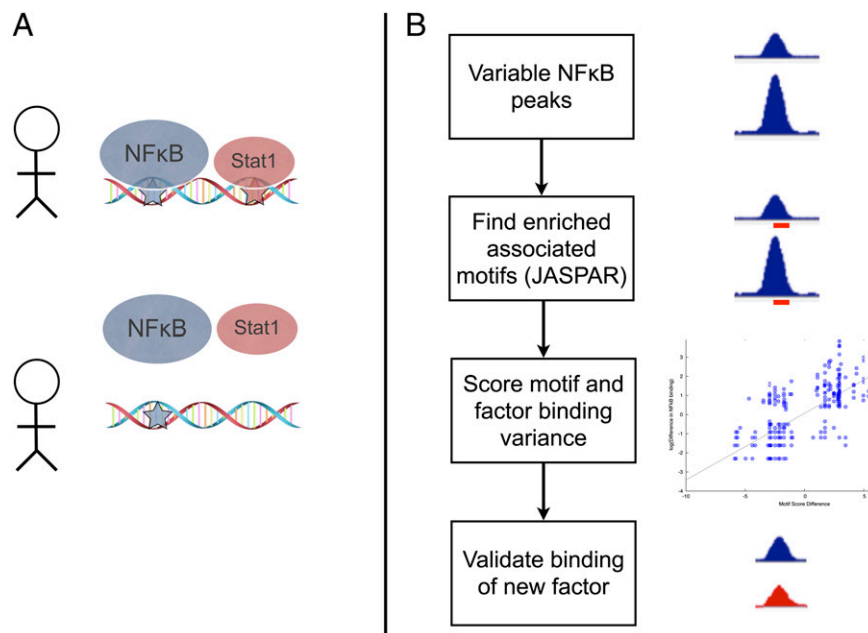

SYSTEMS BIOLOGY

Fig. 1. Overview of the ALPHABIT pipeline. (*A*) In a model of cooperativity, the binding of one factor depends on the binding of another. For example, when a STAT1 motif is present, both STAT1 and NFκB are bound. Loss of the STAT1 motif decreases binding of NFκB, despite the presence of an NFκB motif. (*B*) Associated factor discovery process. Variable motifs are searched for in variable NFκB binding peaks and the difference in "motif score" (i.e., match to consensus) is correlated with difference in NFκB binding. Significant predictions are validated by ChIP-Seq and subject to subsequent analysis.

(Fig. 2*A*). Using a pairwise correlation method, we found 78 motifs where at least 25 comparisons could be made (*SI Appendix*, Fig. S1), where each comparison consists of a motif that is polymorphic between two individuals in a variable NFκB binding region. In this way, we found that five motifs (those for E2A, STAT1, IRF2, EBF1, and Myf), as well as the positive control, NFκB showed significant correlations with NFκB binding after Bonferroni multiple hypothesis correction (*P* < 6.4E-4) (Table 1 and *SI Appendix*, Table S1 and Figs. S1 and S2). The correlation

for one such motif (EBF1) is shown in Fig. 2*B*. Of the five motifs, three (E2A, STAT1, and IRF2) had been suggested to work with NFκB previously at one or a few loci, and the other two (EBF and Myf) have not. For each of these five motifs, a stronger match to the consensus positively correlates with an increase in NFκB binding, suggesting a cooperative rather than antagonistic mechanism of binding. Other factors exhibited negative correlations, but these were below our significance threshold (*SI Appendix*, Table S1).
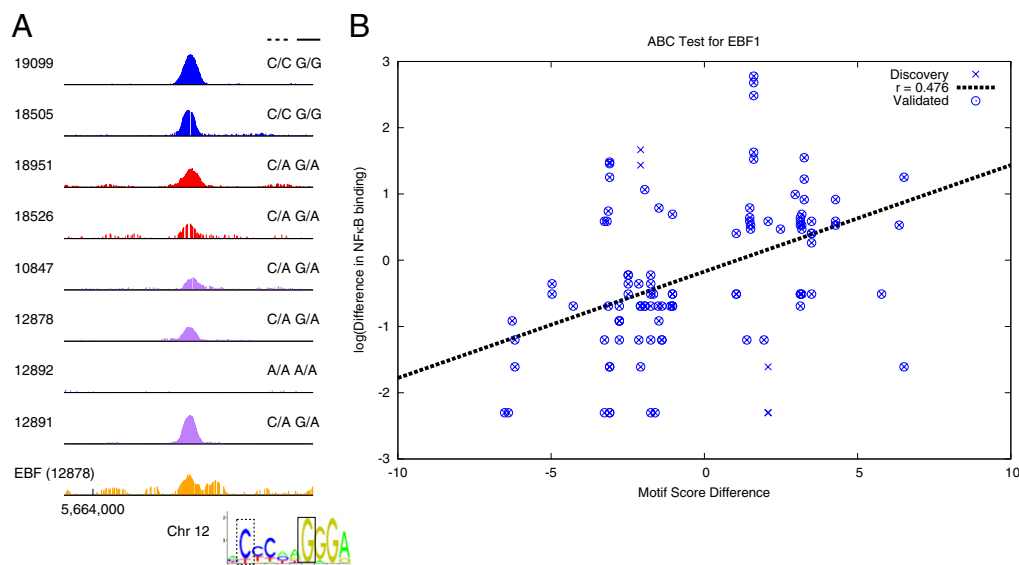

Fig. 2. Output of ALPHABIT analysis. The ALPHABIT pipeline identifies variable motifs that are predictive of NFκB binding. (*A*) Covariance of SNPs in an EBF motif with binding of NFκB at a single locus. Colors correspond to populations (red: YRI; blue: CHB/JPT; purple: CEU), as in ref. 8. Binding of EBF in this region is also validated in GM12878 by ChIP-Seq (orange). The EBF motif is shown below with variable residues highlighted (first genotype in dashed lines, second in solid lines). (*B*) Analysis across all binding sites shows correlation between differences in EBF motifs (quantified as "motif scores") and differences in NFκB binding signal.

Table 1. List of coassociated factors

Motif	De novo pipeline			ChIP-seq validation	Validated loci		
	Loci (comparisons)	Correlation	*P* value	Overall overlap	Loci (comparisons)	Correlation	*P* value
NFκB	90 (370)	0.7544	2.93E-69	N/A	N/A	N/A	N/A
EBF1	44 (108)	0.4756	1.98E-07	55.01%	43 (106)	0.5539	1.29E-09
E2A	9 (29)	0.6426	1.71E-04	30.95%	8 (28)	0.6382	2.58E-04
Myf	62 (170)	0.2779	3.15E-04	ND	ND	ND	ND
STAT1	16 (72)	0.4109	3.36E-04	27.98%	11 (55)	0.3542	7.97E-03
IRF2	58 (111)	0.3507	3.72E-04	28.87%	22 (52)	0.3700	3.92E-03
	. . .				. . .		
ZNF143	38 (101)	0.2172	0.0291	41.24%	32 (78)	0.1413	0.2173
CTCF	63 (154)	0.0778	0.3370	25.8%	55 (130)	−0.0329	0.7105

Factors whose motifs are correlated with NFκB binding, along with the two negative controls, ZNF143 and CTCF. On the left, the significant factors from the discovery stage, along with the negative controls, are shown. A full list can be found in *SI Appendix*, Table S1. Binding was measured for six factors to determine overlap with NFκB binding sites. The correlations are recalculated, restricting the analysis to only validated binding sites (shown on right). N/A, overlap of the same factor is not applicable; ND, no data.

To rule out any biases in the ALPHABIT pipeline toward particular TFs, we performed permutation tests between the motifs and peaks. We found that the pairwise analysis was not particularly biased, as the top results were significant after permutation testing (*SI Appendix*, Table S1). Additionally, these correlations were not an artifact of the particular stringency used in the motif search: we investigated a range of motif cut-off scores and found no dependence of these correlations on the cut-off chosen (*SI Appendix*, Fig. S3).

ChIP-Seq Reveals That the Covariant Motifs Are Bona Fide TF Binding Sites. We then tested whether factors predicted to interact with NFκB with the highest confidence also bound near the NFκB sites using ChIP-Seq, to rule out false-positives resulting from noise in motif discovery. We examined four of the five highest confidence factors (EBF1, E2A, STAT1, IRF2), which are expressed in the GM12878 lymphoblastoid cell line in which NFκB was mapped.

Additionally, because we did not find a significant correlation between the motifs of ZNF143 or CTCF and NFκB binding according to the ABC test (Table 1 and *SI Appendix*, Fig. S2), we examined the binding profiles of these factors as negative controls. In two independent biological replicate experiments, we performed ChIP-Seq for each factor using the same conditions used to map NFκB [TNFα-treated GM12878 cells (8)], along with immunoprecipitation by nonspecific IgG as the control. We then scored the peaks using the PeakSeq algorithm (11) and compared them with the NFκB binding profile. We found that for each factor there was extensive coassociation of binding across all NFκB sites (Table 1). Importantly, the binding of each factor in the NFκB variable regions strongly correlated with the presence of the variable motif; that is, NFκB regions with variable (non-NFκB) motifs were also bound by that motif's factor (Fig. 2A). For example, 43 of 44 NFκB binding regions with a variable EBF1 motif were bound by EBF1 in GM12878 cells (Fig. 2B). Similarly, 11 of 16, 8 of 9, and
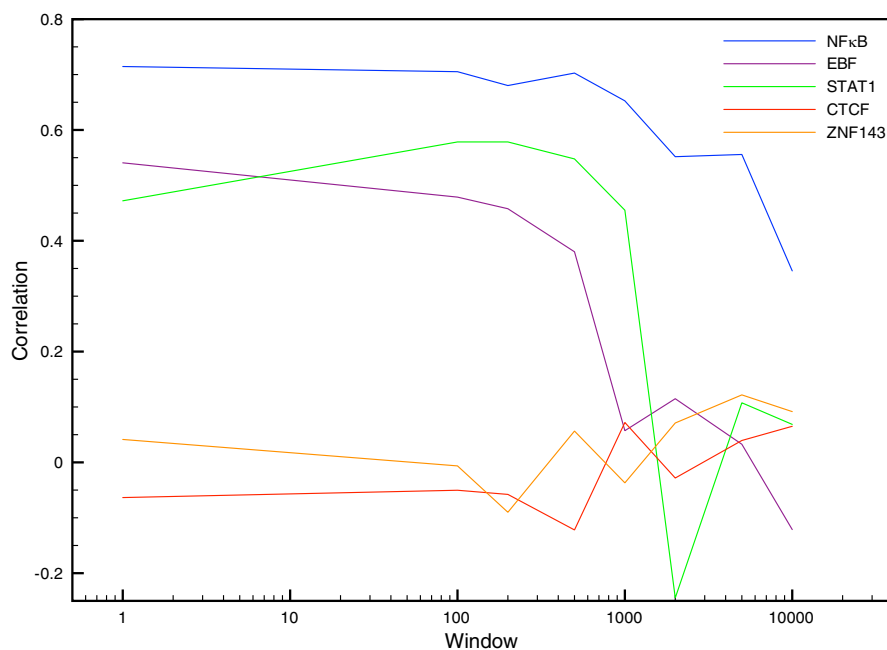

Fig. 3. Distance dependence of cooperative interactions: The correlation between motif PWM match and NFκB binding strength drops when analyzing larger windows around the NFκB binding peak for motifs in validated binding sites. The number of sites was too low for distance dependence analysis of E2A motifs. As expected, proximal motifs are more predictive of NFκB binding than distal motifs.
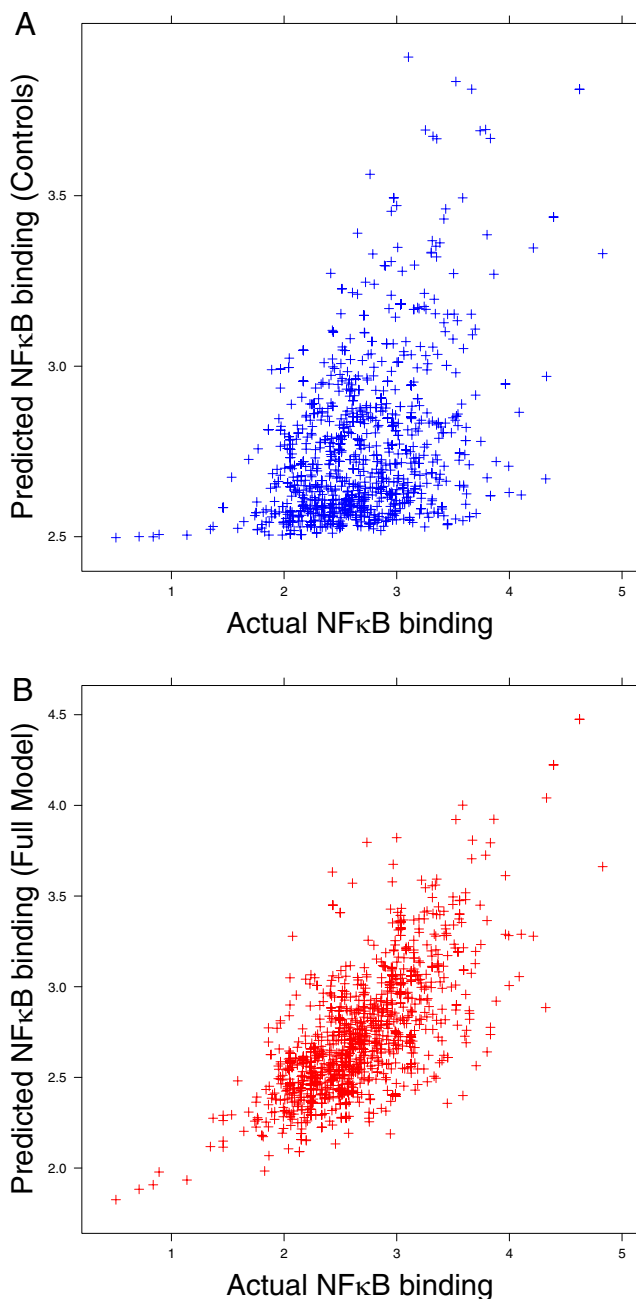
Fig. 4. Correlation of coassociated factors with NFκB binding. NFκB binding ratios (signal to background) can be predicted by fitting linear models of combinations of the control factors (*A*) and the "full" model (the control factors combined with four coassociated factors, EBF, STAT1, E2A, IRF2) (*B*). Although all signals are expected to be correlated because of open chromatin regions, a model fit using the coassociated factors (STAT1, EBF, E2A, and IRF2; *r* = 0.703) (*B*) is significantly more predictive of NFκB binding than one fit using the negative controls (ZNF143 and CTCF; *r* = 0.444) (*A*).

57 of 58 sites with variable motifs were bound by STAT1, E2A, and IRF2, respectively. For each of these factors, a significant correlation is maintained between motifs and binding, when restricting the analysis to only validated binding sites (Table 1). Neither ZNF143 nor CTCF motifs were found to exhibit coassociation with NFκB binding at the validated binding sites of the respective factors (Table 1), and ChIP-Seq of these factors revealed only modest overlap with NFκB binding sites. Thus, the ABC test has high specificity in identifying factors that associate with NFκB.

One concern is that the presence of more than one variable motif in a single binding peak may confound the identification of coassociated factors. For example, because of partial motif similarity, some predicted EBF1 motifs might overlap predicted NFκB motifs; if this occurs, variations that we predict affect the EBF1 motif may actually be affecting an overlapping NFκB motif. Individual inspection of each of the variable motif sites that we identified revealed that no overlap of motifs with either NFκB or one another occurred (*SI Appendix*, Table S2). Thus, we conclude that the associations that we identified using the ABC test do occur with these specific factors in the condition tested.

Covarying Motifs of Interacting Factors Lie Relatively Close to NFκB Binding Peak. We also examined the distance in which functional coassociation between factors can be detected using the ABC test. We determined the motif-NFκB binding correlations, including motifs in various sized windows (in base pairs) around the binding region. For each of NFκB, EBF, and STAT1, we found significant correlations up to 1 to 2 kb, with a significant reduction in correlation past this distance (Fig. 3). For E2A, the number of variable sites was too low for reliable analysis. These results indicate that most coassociated factors can be detected relatively proximal to NFκB sites using the ABC test.

Increased Binding Variance Can Be Explained by Coassociated Factors. Previously, only 0.7% of the variation in NFκB binding could be explained by a SNP in the NFκB motif, with another 3% accounted for by structural variants. Examination of the cB-SNPs in the motifs of the four factors identified in this study (EBF, E2A, STAT1, and IRF2) can account for an additional 0.8% of the variation (127 variable loci) in NFκB binding. Thus, we doubled the number of sites that can be genetically attributable to differences in NFκB binding.

Combinations of TF Binding at NFκB Targets. We used ChIP-Seq data and multivariate linear models to examine which of the factors are preferentially associated with NFκB and whether different combinations of factors are coassociated. In total, 39,818, 22,404, 20,257, and 20,898 of EBF1, E2A, STAT1, and IRF2 sites (both variable and nonvariable) overlap with 72,379 NFκB binding sites. Sites that were not identified as functioning with NFκB using the ABC test—29,853 ZNF143 sites and 18,679 CTCF sites—also overlap with NFκB, perhaps because certain chromosomal regions are more accessible to binding and are bound nonspecifically by TFs.

The associations of the factors predicted by the ALPHABIT pipeline exhibit a higher correlation with NFκB binding strength than that of the negative controls. Considering only the strongest NFκB binding regions (15,931 regions of q-value < 1e-12), NFκB binding strength is more effectively predicted by a multivariate model including STAT1, EBF, E2A, and IRF2 (the "full" model; *r* = 0.70) compared with a model of the control factors, the "null" model, (ZNF143 and CTCF) alone (*r* = 0.44) (Fig. 4), determined by ANOVA (*P* = 1.7e-121, $F_{4,1223}$ = 180.36). In the full model, each of STAT1, EBF, E2A, and IRF2 significantly contribute to the fit of the model (*P* = 1.85e-37, 5.27e-23, 1.31e-06, and 2.92e-11, respectively), whereas neither ZNF143 nor CTCF significantly contribute (*P* = 0.315 and 0.548, respectively). ANOVA results for each model are provided in *SI Appendix*, Table S3. Increased binding of NFκB is predicted by increased binding of these factors, suggesting a cooperative binding mechanism.

We also examined the relationship between NFκB and the predicted factors using 9,395 gene-expression experiments from the Gene Expression Omnibus (GEO). Over all experiments, we observed a significantly higher correlation between NFκB expression and expression of IRF2, STAT1, and EBF1 (0.34–0.39), compared with ZNF143 and CTCF (0.10 and 0.13) (Fig. 5 and *SI Appendix*, Table S4). These results demonstrate that the factors discovered by the ALPHABIT pipeline are significantly coregulated.
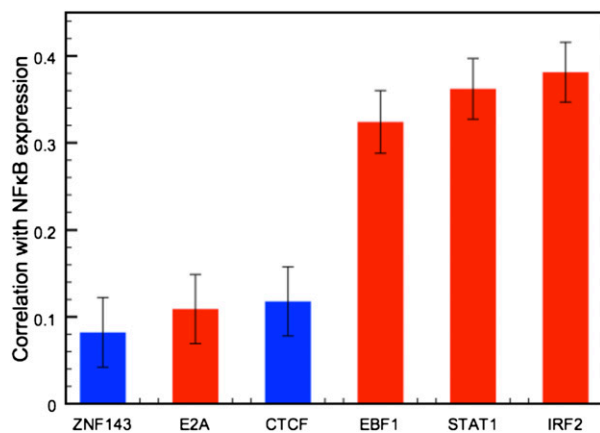
Fig. 5. Coexpression of predicted coassociated factors. Expression of three of the predicted coassociated TFs (IRF2, STAT1, and EBF) exhibit correlation (0.398, 0.379, 0.341, respectively) with NFκB expression in a large, independent dataset (9,395 gene-expression experiments from the GEO). ZNF143, CTCF, and E2A exhibit a much lower correlation (0.102, 0.128, and 0.137, respectively).

Different Combination of Factors Regulate Different Processes. Finally, we examined the functions of the genes regulated by NFκB and each cooperative factor. Using the GREAT algorithm (12), we found functional enrichments of the NFκB coassociated regions with specific gene functions, compared with the background of all NFκB binding sites. We find significant enrichments for various GO terms for each set of overlapping sites of the cooperative factors, but not those of the control factors (Table 2). Interestingly, these functional annotations are different between the cooperative factors, suggesting a distinct functional role for each combination of factors. For example, genes where NFκB and E2A overlap are enriched for genes regulating B-cell homeostasis, but genes targeted by NFκB and EBF1 are enriched for genes involved in melatonin biosynthesis. Interestingly, STAT1 and IRF2, which share an enrichment for "translational elongation," are known themselves to interact, as determined by STRING (13). Thus, we further investigated the combinatorial interaction between these factors: although the subdivisions of biological processes specific to STAT1 and IRF2 varied, none were below the significance threshold (*SI Appendix,* Table S5). Overall, these results indicate a distinct functional role for each combination of factors, as suggested previously with other ChIP data (14, 15).

Discussion

Identification of TF coassociation at specific targets is a significant challenge. Here we present the ALPHABIT pipeline to detect TF associations using ChIP-Seq data obtained from multiple indi-

viduals. We apply this method to binding variation data and identify TFs that are coassociated with NFκB. One important feature of the ABC test and ALPHABIT pipeline is that it not only identifies factors that work together, but also identifies the chromosomal sites at which they operate. Importantly, it also provides functional information for their coassociation, as variance of the motif of one factor is correlated with binding of the other (NFκB). In addition, the binding of the putative coassociated factors is significantly predictive of NFκB binding, suggesting a cooperative mechanism. In this respect, the ABC test/ALPHABIT pipeline is different from other methods that only measure coassociation between factors (6) or protein–protein interactions, but do not ascribe a functional relationship.

Indeed, some of the factors predicted by the ALPHABIT pipeline, including STAT1, E2A, IRF2, are known to interact with NFκB, as determined by STRING (13). Additionally, the top factors we identify exhibit positive correlations between their motifs and NFκB binding: other factors in our analysis are negatively correlated, suggesting a potential antagonistic interaction. Further validation of these results would provide insight into mechanisms of inhibitory combinations of TFs.

Another important advantage of the ABC test/ALPHABIT pipeline is that association can be detected at a limited number of loci because identification occurs by comparing results among different individuals rather than from multiple loci. This finding is important because many factors can facilitate binding, but often at only a limited number of chromosomal sites (9). For E2A we observed a signal from only nine loci, many of which were validated by ChIP-Seq. This low number of coassociated sites could explain its lack of correlation with NFκB expression: E2A may contribute to NFκB binding at a small subset of sites, suggesting localized cooperativity rather than a global coassociation. One current limitation of the ABC test/ALPHABIT pipeline is that data from multiple individuals needs to be present. However, as the cost of DNA sequencing is rapidly decreasing, such data will be easily attainable, and this method can be further applied to discover new TF interactions.

Our results further revealed that significant coassociations are evident within 1 to 2 kb of the binding site of a factor of interest. Although coassociations at more distal sites are likely to be present, they will also be more difficult to detect because of the presence of many other nonfunctional but related motifs that lie in these regions, or alternatively, the effect of more distal sites on binding may be weaker than the proximal sites.

Presently, our analyses have been based on simple linear correlations. However, most biological processes are more complex, and thus a thermodynamic model may be better suited than a linear model to characterize the mechanism of TF binding (16). Such nonlinear complex interactions could include either direct cooperative interactions between TFs, as well as nucleosome-mediated

Table 2. Functional categorization of combinations of TFs

Factor	Term	Raw P value	False-discovery rate q-value	Fold enrichment
STAT1	Translational elongation	9.62E-13	6.90E-09	2.508004
IRF2	Translational elongation	3.33E-08	5.96E-05	2.088436
IRF2	rRNA processing	0.000020028	0.013047312	2.047355
EBF	Melatonin biosynthetic process	6.66E-05	0.005970106	2.658929
EBF	Indolalkylamine biosynthetic process	0.000296952	0.017023681	2.454396
E2A	Inactivation of MAPK activity	8.73E-06	0.003293225	2.308324
E2A	Response to iron ion	2.32E-05	0.006912745	2.545806
E2A	Mammary gland epithelial cell proliferation	2.86E-05	0.007316079	2.454884
E2A	B-cell homeostasis	8.52E-05	0.014885823	2.137337

Overlapping binding sites of NFκB with each factor shows enrichment of different functional annotations. No significant annotations were found for NFκB with either ZNF143 or CTCF. The overlapping binding sites were compared with all NFκB binding sites as a background.

effects (17). The ABC test indicates that factors are coassociated but does not provide insight into the mechanisms by which the associated factors function. The mechanisms could either be direct, such as cooperative binding through protein–protein interactions, or indirect, such as might occur for a factor that alters chromatin state and thereby directly facilitates NFκB binding.

Single nucleotide variants in NFκB binding motifs and structural variants in the motif regions were previously found to account for 0.7% and 3% of the sites with variable binding, respectively (8). By incorporating cB-SNPs, we can account for an additional 0.8% of variable binding sites, which still leaves the majority of the variance unexplained. For the binding sites that contain SNPs, further analysis of other cooperating factors may serve to explain the role of these SNPs in regulatory variation. For the remainder of sites, the variability may be explained by long-range interactions, epigenetic modifications, or noise in ChIP-Seq experiments. Thus, the incorporation of other analyses will be important for a full understanding of differences in TF binding among individuals, which will allow for advances in systems biology, as well as comprehensive annotations of genomic variants.

Finally, our analysis demonstrates the different functions performed by combinations of TFs. For example, NFκB may work with E2A in regulating B-cell homeostasis, but its role with STAT1 may be related to regulating proteins involved in translational elongation. Thus, we elucidate the functional role of various combinations of TFs, further demonstrating the context-specific combinatorial manner in which TFs regulate their targets.

Methods

ABC Test and the ALPHABIT Pipeline. The ABC test was performed using NFκB ChIP-Seq data for 10 individuals, 8 of which have full genome sequences available from the 1,000 Genomes Project Pilot 1 (04/09 release). In this test, we obtained a known motif in the PWM format from the JASPAR database (10) and searched for this motif in every variable NFκB binding peak (restricting to peaks with motif e-values ≤ 1,000). A "motif score" was determined for each k-mer motif for each individual, which we define as the sum over each position of the log-odds score as determined by the PWM. If multiple motifs are found in a binding site, the sum of all of the motif scores is used instead. For a variable motif in a pair of individuals, we compare deviations in this motif score to the fold-difference in NFκB binding. The data for all NFκB binding regions from all pairs of individuals were then combined to generate an overall correlation.

In the ALPHABIT pipeline, the correlations between motif scores and NFκB binding were calculated using the ABC test for each of the 146 motifs in the JASPAR database. Seventy-eight motifs were found to be variable in NFκB regions, where at least 25 comparisons were made: each comparison required both a change the motif score by at least 1, as well as a significant difference in NFκB binding (as determined previously in ref. 8). We determined Pearson correlations between the "motif score" and NFκB binding for each of these motifs independently and applied a Bonferroni correction ($n = 78$) for multiple hypotheses.

To investigate any bias introduced by this pairwise comparison method, we first randomized the motif-peak association by sampling motifs with replacement for each individual. For each iteration, we randomly selected which individual would be compared first (i.e., which individual is used as the numerator for the comparison), and repeated this process 100 times to generate an average correlation. We repeated this process 100 times to generate a background distribution of permuted correlations and estimated the P value by comparing the observed correlation to this normally distributed background.

ChIP-Seq. ChIP-Seq was performed using the procedures of ref. 18, using antibodies for EBF1 (Santa Cruz Biotechnology; sc-137065), E2A (Santa Cruz Biotechnology; sc-349), STAT1 (Santa Cruz Biotechnology; sc-345), IRF2 (Santa Cruz Biotechnology; sc-13042), ZNF143 (Proteintech; 16618–1-AP), and CTCF (Millipore; mp07729). ChIP-Seq for NFκB (Santa Cruz Biotechnology; sc-372) was also repeated to ensure corresponding conditions. Cells were treated with 25 ng/mL TNF-α (eBioscience; #14–8329) for 6 h, as described in refs. 8 and 18. Peaks were scored using the PeakSeq algorithm (11) using a q-value threshold of 0.01. Data are available from the University of California at Santa Cruz.

Statistical Analysis. High-confidence NFκB binding regions determined by an independent ChIP-Seq experiment (15,931 regions of q-value < 1e-12) were intersected with binding sites of each of the other factors, requiring a single base-pair overlap for corresponding binding peaks. Multivariate linear regression was performed on NFκB binding ratios (signal to background) generated by PeakSeq. Significance was assessed by ANOVA of nested models.

For expression correlation analysis, Pearson correlation between NFκB and each of the other factors was determined for a set of 9,395 gene-expression experiments, described in ref. 19. Significance was assessed through bootstrapping. From 20,099 genes, we sampled with replacement 100 times and calculated Pearson correlations. The empirical P value was the proportion of randomly sampled correlations that exhibited higher correlations than the observed correlations. We used R statistical software (version 2.12.1) for all statistical analyses.

We used the GREAT algorithm (version 1.8) (12) to identify functional enrichments of overlapping binding sites. We used the same set of high-confidence NFκB binding regions and considered overlap with only the highest confidence peaks of each of the other factors (q-value < 1e-15 for all factors except EBF, for which more data were available, where we required q-value < 1e-40). For the detailed analysis of STAT1 and IRF2, we considered NFκB peaks where IRF2 was bound, but STAT1 was not, by the same criteria as above, and vice versa.

1. Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EE (2009) Combinatorial binding predicts spatio-temporal cis-regulatory activity. Nature 462(7269):65–70.
2. MacArthur S, et al. (2009) Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. Genome Biol 10(7):R80.
3. Moresco JJ, Carvalho PC, Yates JR, 3rd (2010) Identifying components of protein complexes in C. elegans using co-immunoprecipitation and mass spectrometry. J Proteomics 73:2198–2204.
4. Ravasi T, et al. (2010) An atlas of combinatorial transcriptional regulation in mouse and man. Cell 140:744–752.
5. Beer MA, Tavazoie S (2004) Predicting gene expression from sequence. Cell 117(2):185–198.
6. Borneman AR, et al. (2006) Target hub proteins serve as master regulators of development in yeast. Genes Dev 20:435–448.
7. Yu X, Lin J, Zack DJ, Qian J (2006) Computational analysis of tissue-specific combinatorial gene regulation: Predicting interaction between transcription factors in human tissues. Nucleic Acids Res 34:4925–4936.
8. Kasowski M, et al. (2010) Variation in transcription factor binding among humans. Science 328:232–235.
9. Zheng W, Zhao H, Mancera E, Steinmetz LM, Snyder M (2010) Genetic analysis of variation in transcription factor binding in yeast. Nature 464:1187–1191.
10. Bryne JC, et al. (2008) JASPAR, the open access database of transcription factor-binding profiles: New content and tools in the 2008 update. Nucleic Acids Res 36(Database issue):D102–D106.
11. Rozowsky J, et al. (2009) PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. Nat Biotechnol 27(1):66–75.
12. McLean CY, et al. (2010) GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol 28:495–501.
13. Jensen LJ, et al. (2009) STRING 8—A global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res 37(Database issue):D412–D416.
14. Reed BD, Charos AE, Szekely AM, Weissman SM, Snyder M (2008) Genome-wide occupancy of SREBP1 and its partners NFY and SP1 reveals novel functional roles and combinatorial regulation of distinct classes of genes. PLoS Genet 4: e1000133.
15. Ni L, et al. (2009) Dynamic and complex transcription factor binding during an inducible response in yeast. Genes Dev 23:1351–1363.
16. Gertz J, Siggia ED, Cohen BA (2009) Analysis of combinatorial cis-regulation in synthetic and genomic promoters. Nature 457:215–218.
17. Mirny LA (2009) Nucleosome-mediated cooperativity between transcription factors. Proc Natl Acad Sci 107:22534–22539.
18. Raha D, Hong M, Snyder M (2010) ChIP-Seq: A method for global identification of regulatory elements in the genome. Curr Protoc Mol Biol. Chapter 21: Unit 21.19.1–14.
19. Engreitz JM, Daigle BJ, Jr., Marshall JJ, Altman RB (2010) Independent component analysis: Mining microarray data for fundamental human gene expression modules. J Biomed Inform 43:932–944.