NF-Y coassociates with FOS at promoters, enhancers, repetitive elements, and inactive chromatin regions, and is stereo-positioned with growth-controlling transcription factors

Joseph D. Fleming,[1] Giulio Pavesi,[2] Paolo Benatti,[3] Carol Imbriano,[3] Roberto Mantovani,[2] and Kevin Struhl[1,4]

[1]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA; [2]Dipartimento di BioScienze, Università degli Studi di Milano, 20133 Milano, Italy; [3]Dipartimento di Scienze della Vita, Università di Modena e Reggio Emilia, 41125 Modena, Italy

NF-Y, a trimeric transcription factor (TF) composed of two histone-like subunits (NF-YB and NF-YC) and a sequence-specific subunit (NF-YA), binds to the CCAAT motif, a common promoter element. Genome-wide mapping reveals 5000–15,000 NF-Y binding sites depending on the cell type, with the NF-YA and NF-YB subunits binding asymmetrically with respect to the CCAAT motif. Despite being characterized as a proximal promoter TF, only 25% of NF-Y sites map to promoters. A comparable number of NF-Y sites are located at enhancers, many of which are tissue specific, and nearly half of the NF-Y sites are in select subclasses of HERV LTR repeats. Unlike most TFs, NF-Y can access its target DNA motif in inactive (nonmodified) or polycomb-repressed chromatin domains. Unexpectedly, NF-Y extensively colocalizes with FOS in all genomic contexts, and this often occurs in the absence of JUN and the AP-1 motif. NF-Y also coassociates with a select cluster of growth-controlling and oncogenic TFs, consistent with the abundance of CCAAT motifs in the promoters of genes overexpressed in cancer. Interestingly, NF-Y and several growth-controlling TFs bind in a stereo-specific manner, suggesting a mechanism for cooperative action at promoters and enhancers. Our results indicate that NF-Y is not merely a commonly used proximal promoter TF, but rather performs a more diverse set of biological functions, many of which are likely to involve coassociation with FOS.

[Supplemental material is available for this article.]

Transcriptional regulatory proteins and the RNA polymerase II (Pol II) machinery recruit chromatin-modifying activities to their target loci, thereby determining the genomic pattern of histone modifications and nucleosome occupancy. Activator proteins, functioning combinatorially at distal enhancers and in proximity to core promoters, recruit nucleosome remodeling and histone acetylase complexes, thereby generating nucleosome-depleted regions that nevertheless have peaks of histone acetylation. The Pol II machinery recruits H3K4 histone methylases near the core promoter, and upon transcriptional elongation recruits H3K36 and H3K79 histone methylases to active coding regions. Although less well defined, other DNA-binding proteins and nascent RNA can recruit H3K27 or H3K9 methylases to other genomic regions, resulting in heterochromatic silencing by polycomb complexes (PcG) or HP1, respectively.

As a consequence of the above and other mechanistic relationships between TFs and chromatin-modifying activities, the genome-wide pattern of histone modifications and nucleosome occupancy can be used to classify promoters, enhancers, insulators, and distinct types of heterochromatic regions in a given cell type under a given physiological condition. Using chromatin immunoprecipitation (ChIP), formaldehyde-assisted isolation of regulatory elements (FAIRE), and DNase I hypersensitivity tech-

niques coupled to massively parallel DNA sequencing, such classification of functional genomic regions has been done in several cell lines in the context of ENCODE (The ENCODE Project Consortium 2004, 2007, 2011, 2012). In addition, ENCODE has performed genome-wide mapping of binding sites for ~80 TFs (at the time of writing), most notably in the leukemia cell line K562. These genome-wide maps provide an invaluable resource for uncovering new functional aspects of individual TFs.

NF-Y (also known as CBF, CP1) is a heterotrimeric, DNA-binding TF that is conserved in all eukaryotes (Romier et al. 2003). NF-Y binds specifically to the CCAAT motif (Sinha et al. 1995; Bi et al. 1997) that is frequently found in eukaryotic promoters (Suzuki et al. 2001; Marino-Ramirez et al. 2004). The NF-YB and NF-YC subunits (protein products of *NFYB* and *NFYC*) contain histone-fold domains (HFDs) structurally related to H2B and H2A, respectively (Baxevanis et al. 1995), which mediate formation of a stable histone-like heterodimer (Romier et al. 2003). NF-YA (protein product of *NFYA*) binds to this heterodimer, such that the resulting heterotrimeric complex can bind specifically to the CCAAT motif (Sinha et al. 1995). NF-YA contains the sequence-specific CCAAT recognition domain, and NF-YB and NF-YC also contact DNA through their HFDs (Kim et al. 1996; Sinha et al. 1996; Zemzoumi et al. 1999). All bases of the core pentanucleotide are critical for NF-Y binding, with immediate flanking sequences on both ends also being important for efficient DNA binding in vitro (Hooft van Huijsduijnen et al. 1987; Kim et al. 1990) and in vivo (Testa et al. 2005; Ceribelli et al. 2006, 2008).

[4]Corresponding author
E-mail kevin@hms.harvard.edu

At many promoters, the CCAAT motif is highly positioned ~80 bp upstream of the transcriptional start site (TSS), in either orientation, suggesting that its location is important for gene expression. In essentially all promoters tested, mutation of the CCAAT motif reduces or eliminates transcriptional activity (Dolfini et al. 2009). In addition, functional inactivation of NF-Y subunits or the use of a dominant-negative NF-YA mutant indicates that NF-Y binding is important for the pattern of histone modifications at promoters (for review, see Dolfini et al. 2012). Interestingly, bioinformatic studies comparing gene expression patterns in tumors vs. normal tissues indicate that NF-Y sites are highly enriched in promoters of genes overexpressed in tumors (Rhodes et al. 2005; Sinha et al. 2008; Goodarzi et al. 2009), particularly in the most aggressive cohorts. The importance of NF-Y is further underscored by the early embryonic lethality of an NF-YA mouse knockout model due to defects in cell proliferation and extensive apoptosis (Bhattacharya et al. 2003).

Here we describe the genome-wide analysis of NF-Y binding in three tumor cell lines. Using data generated by ENCODE, we analyze the bound loci with respect to chromatin states and binding by 78 chromatin-associating factors. Our results uncover many new and unexpected aspects of NF-Y biology.

Results and Discussion

Unbiased genome-wide identification of NF-Y binding sites

We performed ChIP with anti-NF-YA and anti-NF-YB antibodies in three cell types (K562, GM12878, and HeLa S3) followed by deep DNA sequencing. Antibodies (Dolfini et al. 2009) were validated by Western blot and IP-WB, showing that NF-YA and NF-YB are specifically recognized (Supplemental Fig. 1A,B). Immunoprecipitated DNA was validated on known NF-Y targets (Supplemental Fig. 1C,D) and the reproducibility between biological replicates was high (Pearson correlations: >0.8).

Using a stringent cut-off (P-value $< 10^{-9}$), we identify 12655, 7932, and 5457 NF-YB binding sites and 4726, 289, and 3726 NF-YA binding sites in K562, GM12878, and HeLa S3 cells, respectively (Fig. 1A). Applying the de novo motif discovery tool MEME to NF-YB peaks in K562 cells, we identify an NF-Y binding motif (Fig. 1B) that corresponds well to the motif derived from ChIP-chip experiments (Dolfini et al. 2009). These high-confidence binding sites, 83% of which have at least one CCAAT motif within each site (with a mean of 1.7 motifs per site), will be used for subsequent bioinformatic analyses. At lower stringency, we identify 14,772 ($P < 10^{-7}$) and 18,523 ($P < 10^{-5}$) NF-YB sites in K562, 81% and 77% of which, respectively, have CCAAT motifs. NF-YB sites with relatively high P-values in the range from 10^{-5} to 10^{-7} contain CCAAT motifs at a rate of ~60%, whereas the genomic background is ~5% for similarly sized regions (Supplemental Fig. 2A). Based on these observations and a peak saturation analysis (Supplemental Fig. 2B), we estimate that there are an additional ~4000 low-affinity NF-Y binding sites in the genome of K562 cells.

The apparently higher number of NF-YB sites with respect to NF-YA sites could be due to target loci bound only by NF-YB. In this regard, in nuclei, NF-YB is more abundant than NF-YA, and NF-YB is present in certain post-mitotic cells whereas NF-YA is not detected (Dolfini et al. 2012). However, the NF-YA and NF-YB data sets are highly correlated (Pearson correlation ~0.7) (Fig. 1C), and quantitative PCR analysis of individual sites reveal threefold higher enrichments for NF-YB than for NF-YA. Furthermore,

analysis of 21 NF-YB sites that appear to lack NF-YA show low occupancy of NF-YB, such that an NF-YA peak was below the detection limit (Supplemental Fig. 3A,B). These results indicate that the NF-YB antibody is more "immuno-efficient" than the NF-YA antibody, and that there are few, if any, genomic sites that are bound by NF-YB, but not NF-YA. For this reason, we will use the NF-YB data set to define NF-Y binding sites in subsequent analyses.

Approximately 39% of NF-Y sites are occupied in at least two cell types, whereas the remaining are cell-type specific (Fig. 1D). In accord with this observation, examination of 14 NF-Y target genes identified previously in different cell lines (Ceribelli et al. 2008) revealed that 13 are bound in K562 and eight are bound in Hela S3. We validated the cell-type specificity of a small number of these loci by ChIP-qPCR (Fig. 1E).

Asymmetric binding of NF-YA and NF-YB to the CCAAT motif

Linking the high-resolution positioning data of NF-Y subunits to the CCAAT motif, we confirm that NF-YA binds directly over the CCAAT sequence (Fig. 2A). Interestingly, the NF-Y complex is asymmetric, with NF-YB binding ~15 bp downstream from the CCAAT motif, as defined by the CCAAT strand (Fig. 2A). This asymmetry fits extremely well with the available biochemical knowledge of NF-Y/DNA contacts (Dolfini et al. 2009) and with the crystal structure of trimer interactions with DNA (Romier et al. 2003; Nardini et al. 2013).

NF-Y targets cell signaling, DNA repair, cell-cycle, metabolic, and gene expression genes

GREAT gene ontology analysis of NF-Y bound loci from K562, GM12878, and HeLa S3 reveals a strong enrichment of genes involved in cell-signaling pathways (ITGA2B, MAPK12, MAPK13), cell cycle (G_2/M checkpoints, regulation of DNA replication), DNA repair (homologous recombination and base excision), and metabolism (cholesterol biosynthesis, polyamines) (Supplemental Fig. 4A,B). The connection to the cell cycle and metabolism are in line with previous findings, and further stress the central role of NF-Y in growth-controlling decisions.

In addition, just below our fold-enrichment cutoff, we found a preponderance of GO terms associated with gene expression, in multiple cell lines, that were highly significant. Upon further analysis, it was apparent that NF-Y significantly targeted genes involved in transcription, mRNA splicing, mRNA editing, mRNA 3'-end processing, and mRNA transport. Included is a large and diverse set of TFs, including the NF-Y genes themselves, members of the transcriptional machinery, and coactivators and corepressors (Supplemental Fig. 5A,B; Supplemental Data Tables 10,11,12). Thus, NF-Y appears to be a regulator of gene expression regulators.

In a separate analysis (IPA, Ingenuity) of signaling pathways, we found that NF-Y preferentially associates with genes involved in the interrelated TP53 (protein product of *TP53*, also known as p53) and TNF-related apoptosis-inducing ligand (TRAIL) pathways (Supplemental Fig. 5C,D). This observation reinforces the notion of a direct and indirect NF-Y/TP53 interplay, with opposing functional consequences depending on the TP53 status of the cell, i.e., proliferation or apoptosis (Imbriano et al. 2012). In addition, it is consistent with anecdotal evidence about the role of NF-Y in apoptosis (Morachis et al. 2010; Hughes et al. 2011), which helps explain the phenotypes of NF-YA overexpression and inactivation experiments (Gatta and Mantovani 2011).
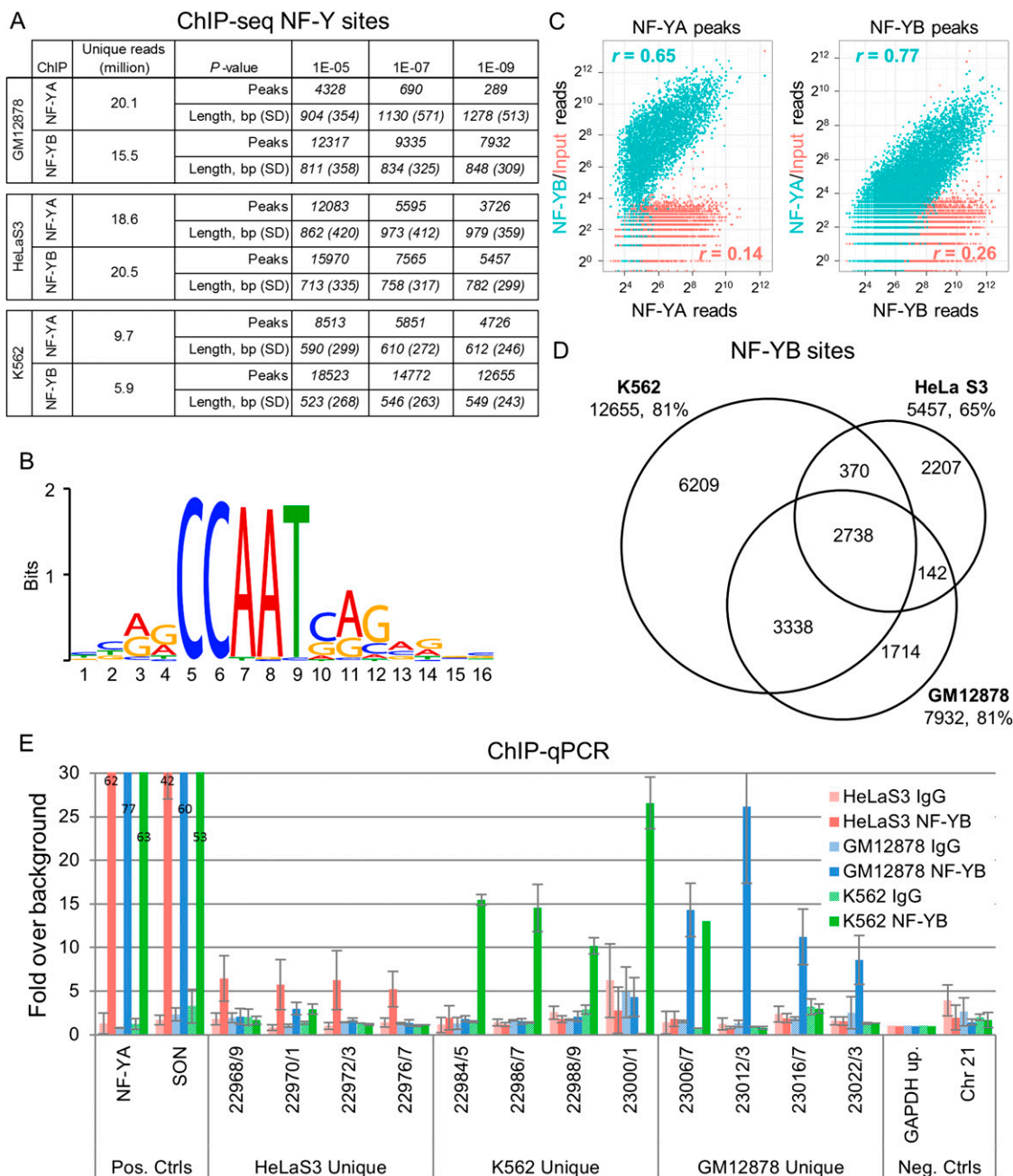
Figure 1. ChIP-seq of two components of the NF-Y complex in three cell types. (*A*) MACS peak analysis indicating peak numbers, mean peak lengths, and standard deviations at three different *P*-value thresholds for NF-YA and NF-YB ChIP-seq data sets in GM12878, HeLa S3, and K562. (*B*) Identification of the NF-Y DNA-binding site motif de novo from 12,655 K562 NF-YB peaks depicted as a sequence logo (Schneider and Stephens 1990). (*C*) Scatter plots of NF-YA, NF-YB, and input read counts at NF-YA or NF-YB sites in K562 showing correlation between data sets. (Blue shading) Correlation amongst NF-YA and NF-YB. (Orange shading) NF-YA or NF-YB correlation to input. (*D*) Venn diagrams depicting the overlap between NF-Y peak populations in GM12878, HeLa S3, and K562. Integers represent peak numbers called at the 10^{-9} *P*-value threshold. The percentages of peaks with CCAAT motifs are indicated (%). (*E*) ChIP-qPCR validation of NF-YB peaks unique to each cell type. (Error bars) Standard deviation of three biological replicates. ''Pos. Ctrls'' are loci known to be bound by NF-Y. ''Neg Ctrls'' are loci known to be devoid of NF-Y. Data represents a fold over background measurement compared with a non-NF-Y bound region (''GAPDH up''). (Solid and striped bars) ChIPs performed with NF-YB specific antibody and nonspecific rabbit IgG, respectively.

NF-Y binds to a diverse set of genomic features including nongenic regions

We annotated the NF-Y bound regions in K562 to RefSeq genes, TSSs, maps of histone modifications and nucleosome-depleted regions, and RNA levels. Unexpectedly, ~25% of the NF-Y binding sites are not situated near RefSeq promoters, or the following types of genic regions: lncRNAs (Khalil et al. 2009); miRBASE (Kozomara and Griffiths-Jones 2011); UCSC RNA genes (Fujita et al. 2011); NONCODE (He et al. 2008); loci bound by Pol II or Pol III (Moqtaderi
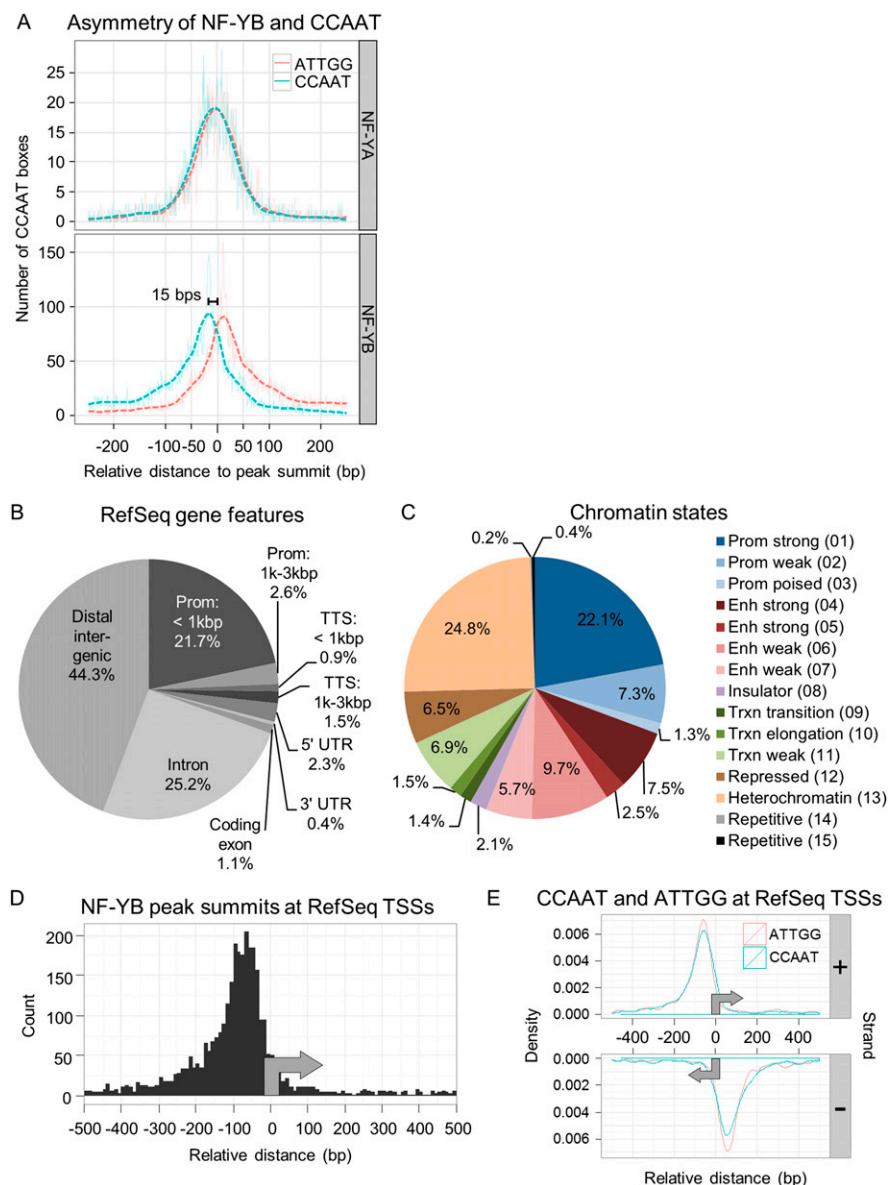
Figure 2. Annotation of NF-Y peaks to genomic features. (*A*) Kernel density estimate of the distribution of the 5′-CCAAT-3′ and 5′-ATTGG-3′ sequences under NF-YA and NF-YB peaks in relation to the peak summit centered at 0 bp. Only the position of the best matching CCAAT motif within 100 bp of the peak summit is considered and plotted. (Solid and dashed lines) Raw and Gaussian smoothed data, respectively. (*B*) Annotation of K562 NF-YB sites to RefSeq gene features. (*C*) As in *B*, except chromatin state maps are used. (Prom) promoter; (enh) enhancer; (trxn) transcription. Numbering is from the chromatin state maps of Ernst et al. (2011). (*D*) Frequency distribution of K562 NF-YB peak summits at RefSeq TSSs showing a preferential location between −50 and −100 bp upstream of the TSS. (*E*) Gaussian kernel density estimate of the distribution of positive and negative strand 5′-CCAAT-3′ and 5′-ATTGG-3′ sequences at K562 NF-YB-bound RefSeq TSSs. Only the best motif per region is considered. Bandwidth is equal to the standard deviation of the smoothing kernel. (Gray arrows) Direction of transcription.

in addition to proximal promoters, where NF-Y has traditionally been observed.

Only a minority of NF-Y binding sites are located at proximal promoter regions

Although NF-Y is typically described as a factor that binds to proximal promoter regions, only 22% of NF-Y sites are located within 1 kbp upstream of a RefSeq TSS (Fig. 2B; Supplemental Fig. 6). A similar analysis shows ~30% of NF-Y sites are located within chromatin states marked by histone modifications characteristic of promoters (Fig. 2C). This is consistent with our previous analysis of 2% of the human genome (Ceribelli et al. 2008). For such proximal promoter binding sites, a frequency distribution plot of peak summits indicates that NF-Y is highly positioned upstream of the TSS at −40 to −100 bp (Fig. 2D), in line with the position of the CCAAT motif at TSSs (Fig. 2E), in agreement with previously published observations (Dolfini et al. 2009). Though NF-YA and NF-YB bind asymmetrically to the CCAAT motif, the orientation with respect to the TSS is largely irrelevant for transcription, as only a small difference in the frequency of CCAAT and its complement ATTGG are noticed on the same strand (Fig. 2E). More generally, only a third of NF-Y loci (clusters B, K, L, N, P, S, U, V; $n = 4061$) (Fig. 3A) are associated with active promoters, as defined by high levels of di- and trimethylated H3K4, acetylated H3K27 and H3K9, Pol II, and nucleosome depletion (defined by a "valley" of low enrichment of mono-methylated H3K4 at NF-Y summits and a FAIRE signal) (Fig. 3A,B). By comparison, essentially no sites are located within nonmodified chromatin regions, and only a few MYC sites are located within weak enhancer-like regions (low K4me1 signals; Supplemental Fig. 7).

A subset of NF-Y sites is located at tissue-specific enhancers

Although NF-Y is typically described as a proximal promoter factor, binding to enhancers has been described, e.g., the 5′

et al. 2010). These sites are not false positives, because the vast majority (88%) have CCAAT motifs, and 46% of them are present in at least one other cell type. Based on the patterns of colocalized histone modifications and Pol II, NF-Y-bound regions in K562 and HeLa S3 reproducibly partition into 20 clusters that can be grouped into five major classes: promoter, enhancer, gene body, PcG repressed, and LTR/nonmodified-chromatin. As discussed below, these results indicate that NF-Y binding is prevalent in tissue-specific enhancers and specific types of repetitive sequences,

upstream regions of the MHC class II genes (Dorn et al. 1988) and the intronic enhancer of the *HOXB4* gene (Gilthorpe et al. 2002). In this regard, all four enhancer chromatin states, as defined by Ernst et al. (2011), are bound by NF-YB (Fig. 2C), totaling 25% of NF-Y peaks in K562. From clustering analysis of histone modifications and Pol II, 12% of all NF-Y sites (clusters E, R, and T; $n = 1525$) have histone modification patterns typical of enhancers: high H3K4me1, low H3K4me2/me3, low Pol II, and only a modest overlap with RefSeq TSSs (Fig. 3A,B). This is also observed with
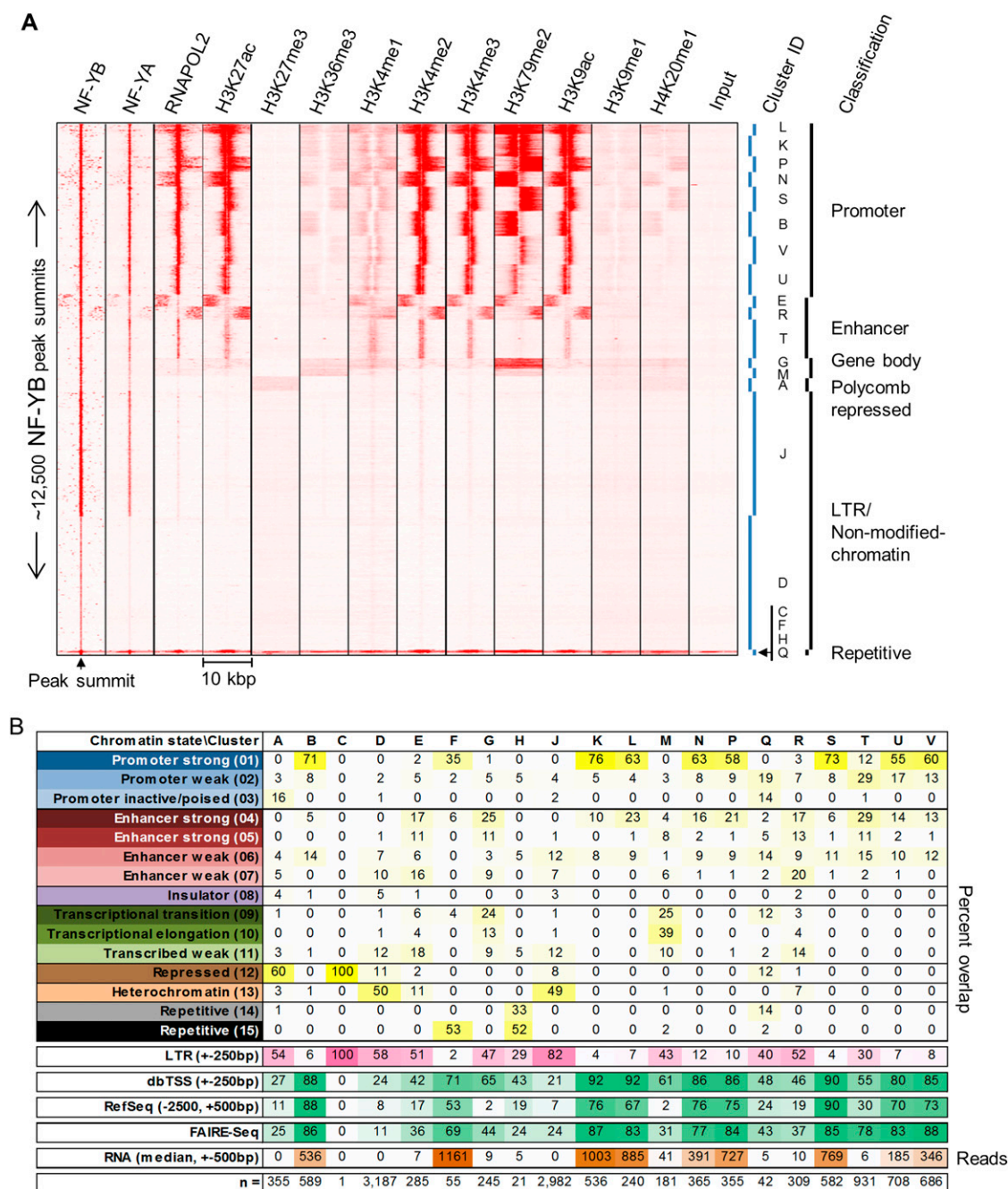
Figure 3. NF-YB bound loci reside within five epigenetic domains. (*A*) K-means clustering of K562 NF-YB loci based on the distribution of histone PTM, RNA Pol II, NF-YB, and NF-YA ChIP-seq reads within a region spanning ±5 kbp from the summit of NF-YB peaks (centered at 0 bp). Clustering was carried out on transformed, rank normalized read counts. Raw read count intensity is depicted in red. The interpretation and classification of clusters into functional categories are shown at *right*. (*B*) NF-YB summits from clusters derived from *A* are annotated to genomic features: chromatin states, LTRs, dbTSS, RefSeq promoters, and FAIRE-seq regions. The percentage of peak summits within each cluster overlapping a specific feature is indicated. Overlap with LTRs is assayed within a window of ±250 bp from the ends of the LTR feature. RefSeq promoters are considered within a window of −2500:+500 bp from the TSS. A direct overlap with FAIRE-seq regions and chromatin states is used. Long poly(A) purified RNA reads were counted within a window of ±500 bp about the NF-YB peak summit, and the median value of that cluster is shown (n = size of cluster in peaks).

Chromatin state\Cluster	A	B	C	D	E	F	G	H	J	K	L	M	N	P	Q	R	S	T	U	V
Promoter strong (01)	0	71	0	0	2	35	1	0	0	76	63	0	63	58	0	3	73	12	55	60
Promoter weak (02)	3	8	0	2	5	2	5	5	4	5	4	3	8	9	19	7	8	29	17	13
Promoter inactive/poised (03)	16	0	0	1	0	0	0	0	2	0	0	0	0	0	14	0	0	1	0	0
Enhancer strong (04)	0	5	0	0	17	6	25	0	0	10	23	4	16	21	2	17	6	29	14	13
Enhancer strong (05)	0	0	0	1	11	0	11	0	1	0	1	8	2	1	5	13	1	11	2	1
Enhancer weak (06)	4	14	0	7	6	0	3	5	12	8	9	1	9	9	14	9	11	15	10	12
Enhancer weak (07)	5	0	0	10	16	0	9	0	7	0	0	6	1	1	2	20	1	2	1	0
Insulator (08)	4	1	0	5	1	0	0	0	3	0	0	0	0	0	2	0	0	0	0	0
Transcriptional transition (09)	1	0	0	1	6	4	24	0	1	0	0	25	0	0	12	3	0	0	0	0
Transcriptional elongation (10)	0	0	0	1	4	0	13	0	1	0	0	39	0	0	0	4	0	0	0	0
Transcribed weak (11)	3	1	0	12	18	0	9	5	12	0	0	10	0	1	2	14	0	0	0	0
Repressed (12)	60	0	100	11	2	0	0	0	8	0	0	0	0	0	12	1	0	0	0	0
Heterochromatin (13)	3	1	0	50	11	0	0	0	49	0	0	1	0	0	0	7	0	0	0	0
Repetitive (14)	1	0	0	0	0	0	0	33	0	0	0	0	0	0	14	0	0	0	0	0
Repetitive (15)	0	0	0	0	0	53	0	52	0	0	0	2	0	0	2	0	0	0	0	0
LTR (+-250bp)	54	6	100	58	51	2	47	29	82	4	7	43	12	10	40	52	4	30	7	8
dbTSS (+-250bp)	27	88	0	24	42	71	65	43	21	92	92	61	86	86	48	46	90	55	80	85
RefSeq (-2500, +500bp)	11	88	0	8	17	53	2	19	7	76	67	2	76	75	24	19	90	30	70	73
FAIRE-Seq	25	86	0	11	36	69	44	24	24	87	83	31	77	84	43	37	85	78	83	88
RNA (median, +-500bp)	0	536	0	0	7	1161	9	5	0	1003	885	41	391	727	5	10	769	6	185	346
n =	355	589	1	3,187	285	55	245	21	2,982	536	240	181	365	355	42	309	582	931	708	686

NF-Y sites in HeLa S3 (Supplemental Fig. 8). The apparent discrepancy in the NF-Y sites in K562 designated as enhancer (25% vs. 12%) is likely due to our more conservative definition of enhancer and wider region used for interpretation. Clusters E and R (Fig. 3A) are exceptional in that they represent NF-Y sites located close to (~2.5 kb), but not within regions of high enrichment for H3K27ac, H3K9ac, H3K4me1/me2/me3 (i.e., strong actively transcribing promoters), unlike all other clusters from the enhancer and promoter groups, where NF-Y is directly within the domains enriched for acetylation and methylation.

Interestingly, cell-type-specific NF-Y sites are enriched for enhancers and are, on average, located further away from TSSs as compared with NF-Y sites common to all cell types (Supplemental Fig. 9A,B). GO analysis of cell type-specific NF-Y loci reveals categories enriched in individual cell types. NFkappaB cascade and regulation of IL12 is enriched in GM12878, a cell type in which NFkappaB is constitutively active (Gubler et al. 1991; Wolf et al. 1991), and HeLa S3 shows enrichment for epidermis morphogenesis and establishment of tissue polarity, commonly associated with cells of epithelial origin (Supplemental Fig. 4B).

Functional inactivation of NF-YA supports a transcriptional role for NF-Y located distally to TSSs

The large number of NF-Y locations at distal enhancers that are functional as defined by the histone modification pattern (Fig. 2C) and previous analysis of individual genes (Dorn et al. 1988; Gilthorpe et al. 2002) strongly suggests that NF-Y binding to distal locations can have functional consequences. To provide additional evidence for this idea, we performed expression array analysis on HeLa S3 cells depleted for NF-YA by lentiviral small hairpin RNA (shRNA) (Supplemental Fig. 10A,B) and correlated these changes to the location of NF-Y (Supplemental Figs. 10C, 11A,B). At a P-value cutoff of 10^{-4}, 84 genes are down-regulated and 252 genes are up-regulated (Supplemental Fig. 10C) upon NF-YA knockdown. Of these, only 11% ($n = 9$) and 39% ($n = 98$) have NF-Y bound to their proximal promoters, respectively. The topmost differentially down- and up-regulated genes both trend toward having a higher percentage of their promoters occupied by NF-Y than non-differentially regulated genes (Supplemental Fig. 11A). Of the 1059 NF-YA peaks in HeLa S3 located within 250 bp of a RefSeq TSS, only 5.2% are differentially regulated at a P-value of 10^{-4} ($n = 55$). The low percentage of differentially regulated genes bound by NF-Y is similar to that found with other TFs (Yang et al. 2006; Strub et al. 2011; Martynova et al. 2012) and could be exacerbated by the incomplete functional inactivation of NF-Y.

We ranked NF-Y sites by the fold change in RNA expression of the nearest associated gene upon NF-YA inactivation. Importantly, for both promoters and distal regions, the topmost differentially down- and up-regulated genes both trend towards having a higher percentage of NF-Y occupancy than nondifferentially regulated genes (Supplemental Fig. 11A). In addition, the most strongly down-regulated genes have NF-Y sites that are much more distal to the TSS, with the median distance being >10 kb; this preference for distal sites is not true for MYC (Supplemental Fig. 11B). This observation suggests the possibility that distal NF-Y sites might be more important for differential regulation. Whatever the relative importance of NF-Y at promoters and enhancers, our results support the idea that NF-Y located at both promoters and enhancers can be important for transcription of neighboring genes.

LTRs are the most prevalent class of NF-Y sites in the human genome

Of all NF-Y binding sites in K562, 40% directly overlap an LTR, the promoter elements of endogenous retroviruses, making LTRs the most prevalent class of NF-Y loci in the human genome (Fig. 4A). NF-Y selectively associates with the MLT1 and LTR12 families of LTRs (Fig. 4B,C). NF-Y does not bind to all LTR families, irrespective of the presence of a CCAAT motif in the consensus sequence. The R66 tandem repeat (which is related to LTR12B) (Benachenhou

et al. 2009a,b), MER51A and MER51E are also associated with NF-Y. In general, there is no significant cell-type specificity in LTR binding (Supplemental Fig. 12).

Most NF-Y bound sites at LTRs lack any detectable histone modifications within 5 kbp of the NF-Y peak summit (Figs. 3A,B, 4D, clusters D and J) . These NF-Y loci appear to be inactive, yet maintain substantial NF-Y occupancy. In contrast, a sizeable minority of LTRs (27% K562; 20% GM12878) (Fig. 4D) are associated with high levels of H3 acetylation and/or H3K4 methylation and appear to be transcriptionally active. This minority class most likely represents NF-Y-bound functional regulatory elements derived from transposable repetitive elements and regulating endogenous genes.

LTRs function as promoter elements of endogenous retroviruses and they can act as regulatory elements for certain host genes (Bourque 2009). NF-Y sites abound in viral LTRs (Graves et al. 1986; Dutta et al. 1990; Faber and Sealy 1990; Greuel et al. 1990; Scheef et al. 2002). The selectivity for the gamma-retrovirus LTR family, and within it for certain members, likely reflects the presence of CCAAT in the original viral LTRs. Thus, our results suggest a strong genetic pressure on their genomic transduced copies to maintain NF-Y binding. This is not unprecedented, as evidenced by the preference of particular TFs for specific repetitive sequences (Bourque et al. 2008; Kunarso et al. 2010). Genetic analysis of the *ERV-9/LTR12* element located 5′ of the globin locus-control region indicates a crucial role of the 14 CCAAT and GATA containing E3 repeats for expression of the β-globin locus (Yu et al. 2005; Pi et al. 2010). Despite this precedent, most NF-Y sites are associated with heterochromatin-like domains and are apparently devoid of any transcriptional signal. As the vast cohort of endogenous LTR proviral sites are under strong control by the host organism and, in most cases, actively repressed (Bourque 2009), we are tempted to speculate that NF-Y plays a role in the repression of these LTRs in somatic tissue and/or in their activation during embryogenesis, where many repetitive elements are demethylated and become expressed (Maksakova et al. 2008).

NF-Y binds CCAAT motifs in nonmodified chromatin domains in vivo, unlike most TFs

The majority of NF-Y sites ($n = 6169$; 49%) are in two similar clusters (D and J, i.e., LTR/nonmodified-chromatin class in Fig. 3A,B) that display no positive or repressive histone modifications tested for, negligible Pol II and polyA RNA, and overlap few open regulatory regions (11%, 25%) and RefSeq TSSs (7%, 11%). An analysis of MYC sites reveals the absence of binding at nonmodified chromatin sites (Supplemental Fig. 7). Interestingly, most of these loci overlap LTRs, 58% and 82%, respectively (Fig. 3B). These NF-Y sites are interesting, as most TFs are believed to not be able to bind to their DNA motifs within closed, transcriptionally inactive chromatin domains.

To further explore this issue, we calculated the percentage of motifs residing within NF-Y peaks within distinct chromatin states, over a range of motif quality scores. Interestingly, and unlike other TFs such as E2Fs and MYC, NF-Y is not excluded from any chromatin state assayed (Fig. 5A,B). At strong and weak promoters, >80% of CCAAT motifs (scores >16) are occupied by NF-Y (Fig. 5A). CCAAT motifs at enhancers and insulators are also well occupied by NF-Y (30%–75%, respectively) (Fig. 5A), although the percent occupancy is lower than at strong promoters, indicating that binding to these genomic regions is more selective. More generally, CCAAT motifs situated within open chromatin regions, as defined by FAIRE, are exceptionally well occupied to near-saturated levels
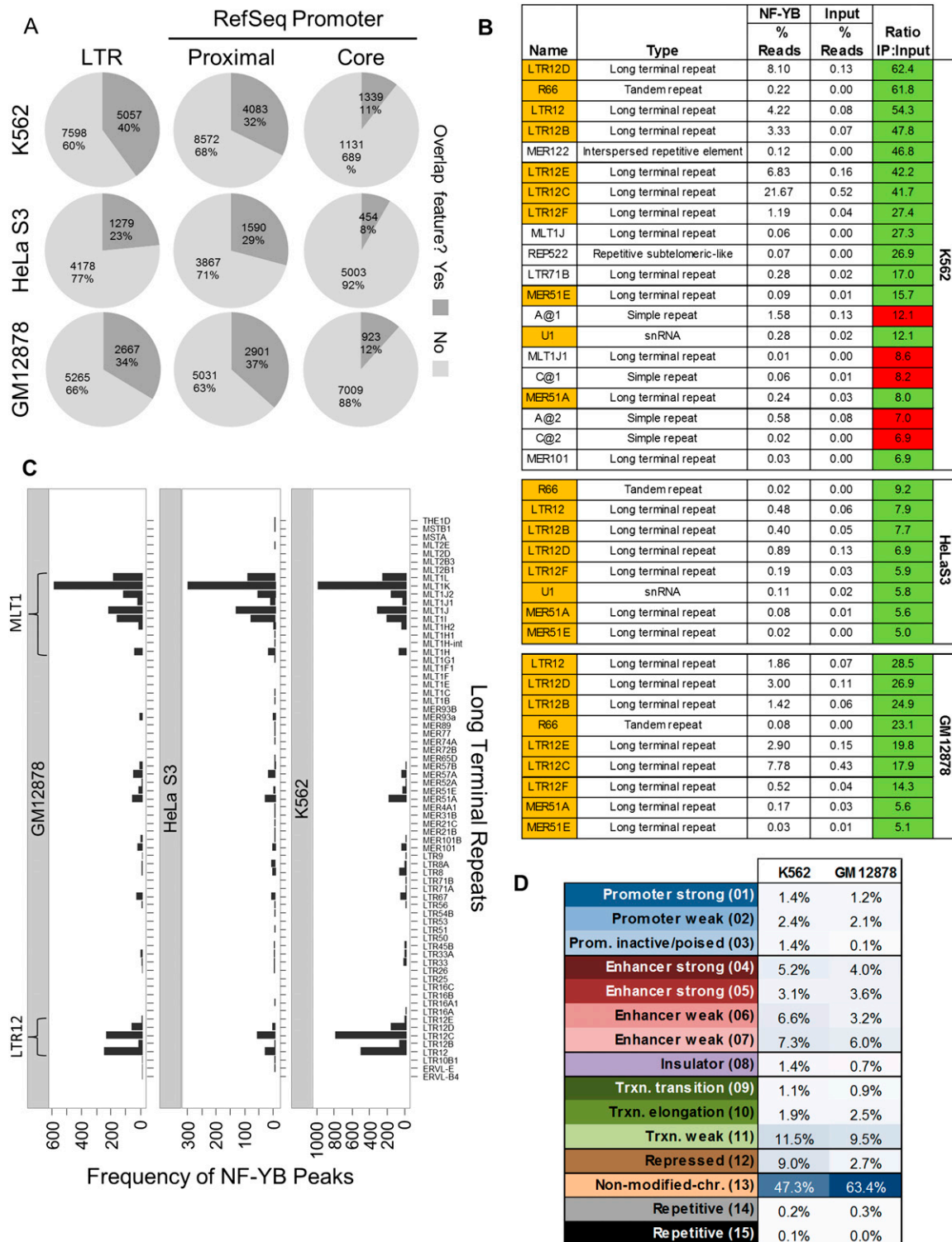
Figure 4. NF-YB binds extensively to long terminal repeats. (*A*) The percentage of all K562 NF-YB peak summits that occupy the indicated feature. Core and proximal promoters are defined as −250:+50 bp and −2500:+500 bp from the TSS of RefSeq promoters, respectively. (*B*) Mapping of ChIP-seq reads from K562, GM12878, and HeLa S3 to Repbase consensus sequences showing an abundance of NF-Y ChIP-seq reads mapping to repetitive elements. Ratios reflect the enrichment of reads in the NF-YB ChIP sample as compared with input. Only Repbase entries with a read ratio ≥5 are shown. Orange shading indicates enriched repeats present in all cell lines. Green and red shading indicate the presence and absence, respectively, of a CCAAT motif match at *P*-value < 10⁻⁴ in the consensus sequence. (*C*) Frequency of overlap between NF-YB peak summits and the genomic locations of LTR families. Only LTR elements that overlap at least one NF-YB summit in each cell line are shown. The two most highly overlapping repeat families are indicated, LTR12 and MLTJ1. (*D*) Distribution of NF-YB bound LTRs from K562 and GM12878 at chromatin states. No chromatin state map is available for HeLa S3.
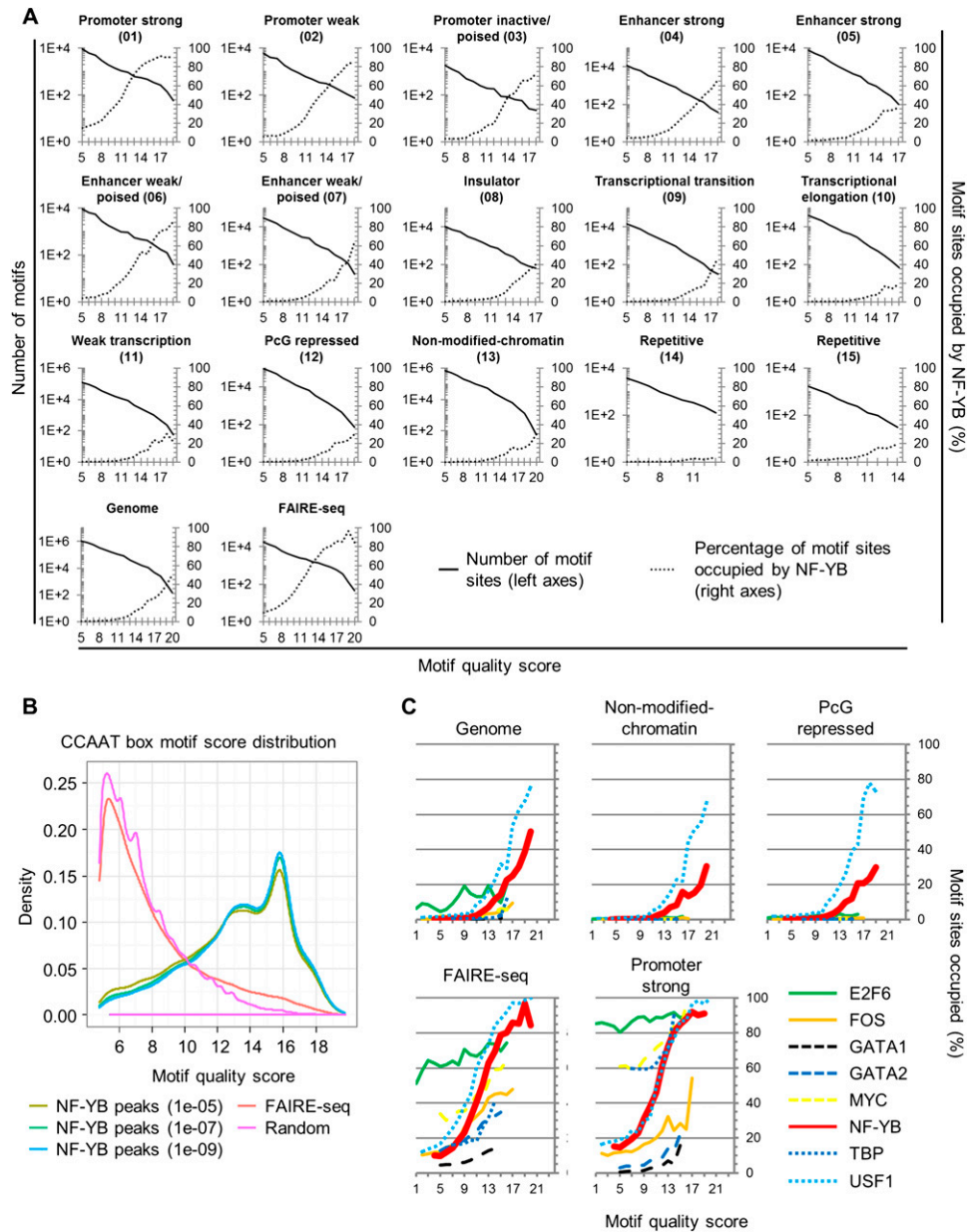
Figure 5. NF-YB can occupy its motif in closed chromatin. (*A*) The percentage of genome-wide computationally discovered CCAAT motifs within each chromatin state, FAIRE-seq regions or the entire genome, that directly overlap NF-YB K562 sites plotted as a function of CCAAT motif quality (*right* axes). Also shown are the numbers of discovered CCAAT motifs as a function of quality (*left* axes). Numbering is derived from Ernst et al. (2011) and kept for consistency. (*B*) Distribution of CCAAT motif quality scores under NF-YB K562 peaks, called at three different *P*-values, a random genomic background sample set of 400 k 500-bp regions and K562 FAIRE-seq regions. (*C*) Similar to *A*, except motifs of different TFs are plotted as a function of motif quality. Only a subset of TFs is shown; see Supplemental Figure 13 for all TFs analyzed.

by NF-Y, with 80% occupancy (Fig. 5A). Interestingly, many CCAAT motifs within the nonmodified chromatin, PcG repressed and transcription elongation states are occupied by NF-Y at a rate of ~20% (Fig. 5A).

To test whether the substantial occupation of CCAAT motifs within nonmodified chromatin and repressed genomic contexts is unique to NF-Y, we performed the same analysis on 22 additional TFs, whose binding sites in K562 cells have been determined by ENCODE. As expected, most TFs show high levels of motif occupancy at nucleosome-depleted regulatory regions at high levels,

comparable to that of NF-Y (Fig. 5C; Supplemental Fig. 13). In contrast, GATA1 and GATA2, thought to be "pioneer" TFs (Magnani et al. 2011; Zaret and Carroll 2011), are highly selective and unable to saturate their motifs that reside within these nucleosome-depleted regulatory regions. However, most TFs lack the ability to occupy even their highest quality motifs within nonmodified and repressed chromatin states. For the 23 factors tested, only USF1, MAFK, and NF-Y can bind to motifs in the context of nucleosomes lacking some of the most common "positive" histone modifications or containing the repressive H3K27me3 mark.

By preventing accessibility to target sites, chromatin is a formidable barrier for binding by most TFs. This creates a dilemma as to how *cis*-regulatory motifs can provide transcriptional competency if they cannot be accessed by their corresponding TFs. There are a small number of "pioneer" factors that can efficiently bind to their DNA motif located within nonmodified, closed chromatin. Once bound, these pioneer TFs can recruit chromatin-modifying activities to generate open chromatin for the subsequent binding of partnering TFs (Magnani et al. 2011; Zaret and Carroll 2011). NF-Y can associate with a CCAAT motif after nucleosome assembly in vitro, and the NF-YB/NF-YC HFD dimer can physically interact with H3/H4 in solution and on DNA (Caretti et al. 1999). Indeed, NF-Y binding is not mutually exclusive with nucleosomes in vitro, giving NF-Y the theoretical functional ability to interact efficiently with chromatin-bound CCAAT motifs in vivo. NF-Y binds to a sizeable number of sites either in functionally "hostile" environments or sites lacking all of the common positive histone modifications. Perhaps the structural features of the HFD heterodimer are instrumental for this. We propose that NF-Y is a type of "pioneer" TF that retains histone-like features while possessing high-sequence specificity, with the ability to access its motif irrespective of the chromatin state.

NF-Y extensively coassociates with FOS, typically at loci lacking an AP-I motif

Unexpectedly, we observe a remarkable coassociation of NF-Y and FOS (Pearson = 0.74) that is only marginally lower than that observed between the NF-Y subunits (Pearson = 0.77) (Fig. 6A). This coassociation is observed over all chromatin states, cluster classes, and genic features, with 45% of NF-Y sites directly overlapping a FOS site and 39% of FOS sites directly overlapping an NF-Y site (Fig. 6B). Similar results are observed in HeLa S3 cells, although the overlap is slightly lower (26% and 16%, respectively). Interestingly, NF-Y does not significantly coassociate with JUN (Pearson = 0.14), a TF that forms heterodimers with FOS (Fig. 6A). NF-Y and FOS sites are located just as close (<50 bp) as that observed between the NF-Y subunits or between FOS and JUN (Fig. 6C). Interestingly, most sites bound by NF-Y and FOS lack detectable AP-1 motifs (Fig. 6D), with the notable exception being sites at LTR loci (see below). In contrast, and as expected, most sites bound by FOS and JUN have an AP-1 motif (Fig. 6D). A representative example of the interplay is shown in Figure 6E.

NF-Y and FOS protein–protein interactions have not been described, but the striking colocalization of these factors suggests the possibility of a direct interaction in the context of chromatin. Alternatively, the coassociation at genomic sites in vivo may not be due to a direct interaction between these factors. In this regard, ChIP-seq analysis reveals that sites bound by JUN N-terminal kinases (JNKs) generally lack AP-1 motifs, but often contain CCAAT motifs (Tiwari et al. 2012). Furthermore, NF-Y is necessary for the JNK association with these genomic sites. We speculate that FOS, together with another protein(s) such as JNK, binds to NF-Y at CCATT motifs in the context of chromatin.

NF-Y coassociates with different TFs depending on genomic context

Given the availability of 78 ChIP-seq data sets in K562 for chromatin associating factors involved in diverse functions, we explored combinatorial genomic interactions of these factors with NF-Y on promoters and enhancers. On a pairwise basis, we observe coassociation between NF-Y and 44 factors at promoters and 50 factors at enhancers ($P < 10^{-10}$, Supplemental Fig. 14A), consistent with the general requirement for multiple TFs to stimulate transcription in mammalian cells. Hierarchical clustering (Supplemental Fig. 14B), and analysis of higher order combinations of factors (Supplemental Fig. 14C), reveals distinct groups of factors at proximal promoters and enhancers. At promoters, the core group of NF-Y coassociating factors includes FOS, CHD2, TBP, Pol II, CCNT2, HMGB2, MYC, and E2F4/-6 (Supplemental Fig. 14B). A comparison of promoters that are or are not bound by NF-Y reveals that only FOS and (to a lesser extent) CHD2 are specifically associated with NF-Y. FOS associates with 59% of promoters bound by NF-Y, but only 8% of promoters not bound by NF-Y. The other factors were common promoter bound TFs (Supplemental Fig. 14C). At enhancers, NF-Y forms a well-defined cluster consisting of FOS, USF1/2, MAX, CHD2, and E2F4 (Supplemental Fig. 14B), a slightly different grouping compared with promoters. FOS and USF1 are highly prevalent, being present, respectively, at 39% and 27% of NF-Y enhancers, and were the most common 2-way overlap at 13% (Supplemental Fig. 14C). A summary of the interactions at promoters and enhancers is shown in Figure 7.

The widespread partnership of NF-Y with a group of TFs (FOS, MYC, and E2Fs) that control cellular proliferation and play important roles in cancer are consistent with the importance of NF-Y for expression of growth-regulating genes. The close association of the E2Fs and NF-Y is consistent with the high enrichment of their motifs at promoters of genes overexpressed in tumors (Rhodes et al. 2005; Sinha et al. 2008; Goodarzi et al. 2009). In addition, apoptosis mediated by overexpression of NF-Y is abolished in cells lacking E2F1 (Gurtner et al. 2010). E2F4 is part of the DREAM complex (Litovchick et al. 2007; Schmit et al. 2007), which binds to the CDE motif, and cooperates with the CCAAT motif to negatively regulate expression of G_2/M-specific genes during the cell cycle (Muller and Engeland 2010; Muller et al. 2012). CCAAT motif and CDE containing G_2/M genes are significantly overexpressed in a model of stepwise transformation of primary fibroblasts (Tabach et al. 2005).

Essentially, all E box binding TFs present in ENCODE are statistically enriched at NF-Y locations (Supplemental Fig. 14A), suggesting a pervasive partnership between CCAAT and E boxes. Interestingly, the number of MYC/NF-Y bound promoters exceeds those with MAX/NF-Y, suggesting that either MYC heterodimerizes with another E box binding partner, or that it binds in an E box-independent manner, possibly directly to NF-Y (Izumi et al. 2001; Ravasi et al. 2010).

At the LTR/nonmodified chromatin class, NF-Y extensively colocalizes only with FOS, USF1, and to a lesser degree, USF2 and SP1, although specific groupings occur upon clustering (Supplemental Fig. 15A). The chromatin at colocalized NF-Y/USF sites is not acetylated, suggesting that USF1 and USF2, in the context of binding with NF-Y, do not behave as barrier elements of acetylated chromatin in intergenic regions that inhibit the spread of heterochromatin (West et al. 2004; Huang et al. 2007; Li et al. 2011). It should also be noted that cluster HL2 ($n = 147$; Supplemental Fig. 15A) is enriched for four members of the CTCF–cohesin insulator complex (CTCF, CTCFL, RAD21, and SMC3) in direct proximity with NF-Y, and a similar small cluster is also observed in the PcG repressed class (data not shown). Interestingly, the NF-Y-bound regions in the LTR/nonmodified chromatin class show a remarkable paucity of known DNA motifs, with the exception of the CCAAT motif and in K562, but not GM12878, the KLF4 motif ($P = 1.6 \times 10^{-10}$; data not shown). De novo motif analysis of the same NF-Y sites again reveals the CCAAT and KLF4 motifs as well as two
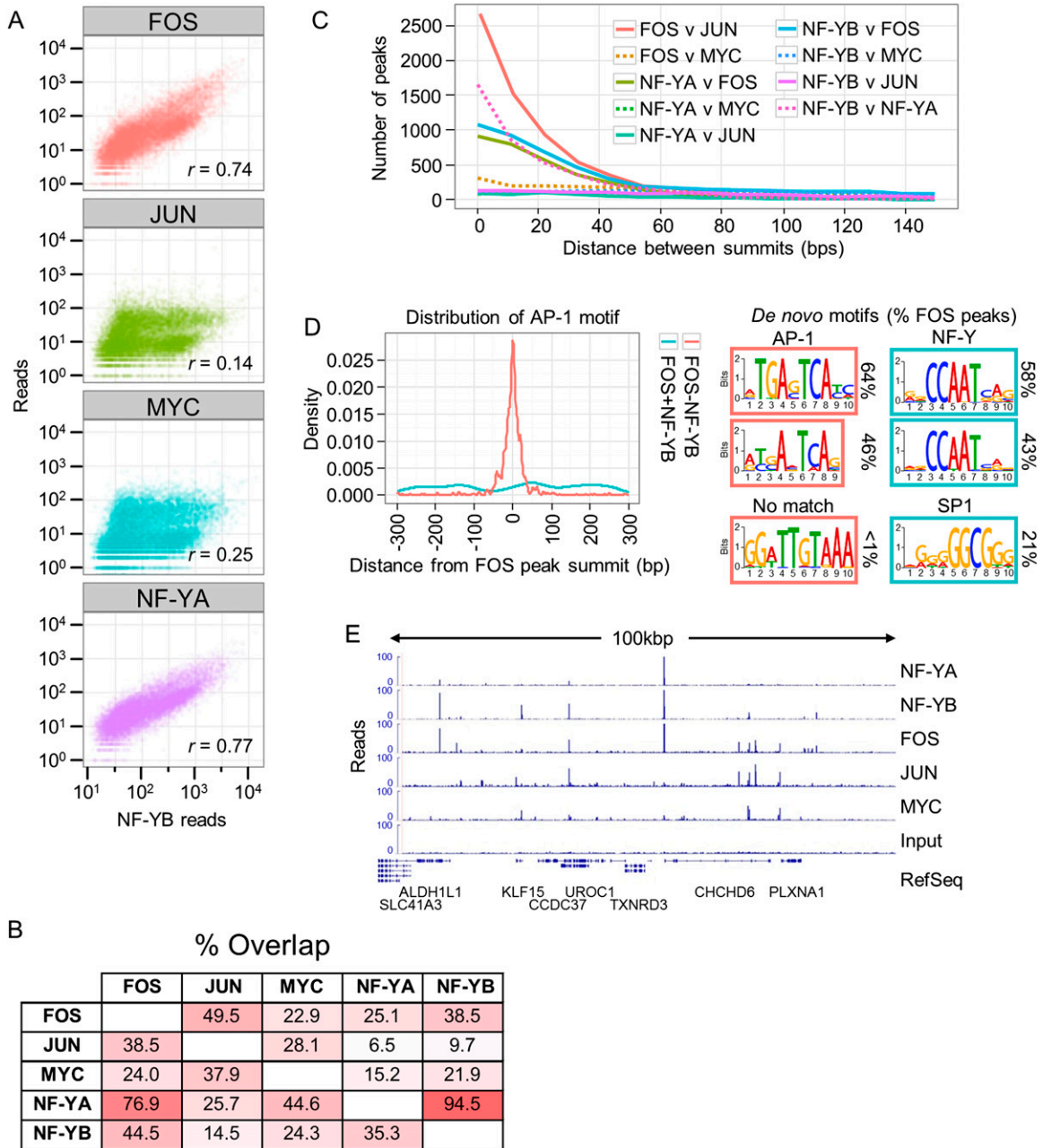
Figure 6. NF-Y and FOS are closely coassociated at loci that lack JUN and the AP-1 motif. (*A*) Correlation between ChIP-seq read counts at NF-YB peak summits, within a window of ±500 bp, between NF-YB and NF-YA, FOS, JUN, or MYC in K562 cells. (*B*) Values represent the percentage of peak populations (*left* row) directly overlapping the peak population of a second factor (*top* column). All binding sites are called at a *P*-value < 10^{-9}. FOS (*n* = 14404); JUN (*n* = 18480); MYC (*n* = 13693); NF-YA (*n* = 4726); NF-YB (*n* = 12655). (*C*) The number of ChIP-seq peaks at the indicated distance between adjacent peak summits is plotted. All peaks were called at a 10^{-9} *P*-value threshold in K562. (*D*) The top 1000 K562 FOS ChIP-seq sites, as ranked by site *P*-value, that directly overlap an NF-YB site ("FOS+NF-YB") and the top 1000 that do not overlap an NF-YB site (10^{-5} *P*-value site list, "FOS-NF-YB") are assayed for the distribution of the AP-1 motif in relation to the FOS peak summit centered at 0 bp. Plotted is the Gaussian kernel density estimate of the AP-1 motif using a bandwidth of 0.5 of the standard deviation of the smoothing kernel. The top three motifs discovered de novo from each FOS peak set, as above, are depicted with the percentage of FOS peaks containing a match to that motif indicated. (*E*) Representative view of a locus on chromosome 3 of the K562 ChIP-seq read counts from NF-YA, NF-YB, FOS, JUN, and MYC ChIPs, with an input control.

unknown motifs (Supplemental Fig. 15B). As KLF4 can act as a transcriptional activator or repressor (Turner and Crossley 1998; van Vliet et al. 2000; Schuierer et al. 2001; Yoon and Yang 2004; Evans et al. 2007; Oishi et al. 2008) and is expressed in K562 cells (Kalra et al. 2011), it may cooperate with NF-Y to repress LTR elements.

NF-Y sites contain positionally biased TFs

To investigate whether there is a specific distance relationship between NF-Y and coassociating factors, we plotted the distribution of the relative position of the TATA element, E box, E2F,
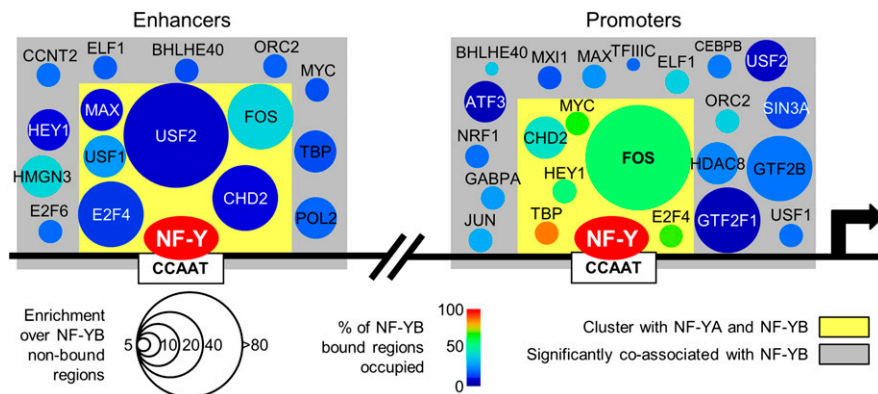
Figure 7. NF-Y coassociates with many factors at promoters and enhancers. Illustration of the factors that significantly associate with NF-YB-bound strong promoters and enhancers. Only those factors that satisfy the following criteria are shown: greater than the median fold enrichment with respect to NF-YB-nonbound regions (enrichment indicated by circle size); greater than the median value of percent occupancy of NF-YB-bound regions (percentage occupied indicated by color); significantly coassociate with NF-Y (gray box, see Supplemental Fig. 14A). Factors enclosed within a yellow box are, additionally, the subset of factors that cluster with NF-YA and NF-YB (see Supplemental Fig. 14B). A black arrow indicates the start of a transcribed region. Two vertical slashes are used to represent being distal to a promoter area.

and AP-1 motifs (termed "predicted") at NF-Y peaks in relation to the position of the best-scoring CCAAT motif, while maintaining strandedness. We then plotted the subset of motifs (termed "verified") that were actually occupied in vivo by the TF of interest. Remarkably, there is an AP-1 motif 10- to 11-bp upstream of CCAAT, which corresponds to verified FOS target sites (Fig. 8A). However, this positioning was only found at NF-Y-bound LTRs, as sites with NF-Y and FOS generally do not contain an AP-1 motif (Fig. 6D). The TATA element (+50), E box ($-12/-11$) and E2F (+6/+7, +31, +55, and +72) motifs are also highly positioned in a CCAAT orientation-specific manner (Fig. 8B). The position of the TATA element is maintained in TBP-bound locations at NF-Y sites. The E2F motif is unusual in that multiple stereo alignments are present and only one, the closest to CCAAT, is maintained at E2F6, but not at E2F4 occupied sites (Fig. 8B; data not shown). The positioning of the E box is only maintained when MAX or USF1 but not MYC loci are considered, suggesting that MYC, when associating with NF-Y, is either not positioned or does not bind DNA directly.

The USF1 observation is interesting because it is one of the few factors that partners with NF-Y in the LTR/nonmodified chromatin class and can bind its motif within a repressive nucleosomal structure. Perhaps the precise positioning may facilitate the cooperation of NF-Y and USF1 to penetrate inactive, non-modified chromatin domains.

Cooperativity mediated by precise spacing between NF-Y and other TFs has been observed at MHC class II promoters, NF-Y/ATF6 sites in ER-stress responsive promoters (Yoshida et al. 2000), and multiple CCAAT motifs in G_2/M promoters (Salsi et al. 2003). Our results greatly extend these findings of precise spacing relationships with NF-Y with its most common TF partners, notably those that play crucial roles in the control of cell proliferation, cell cycle, and metabolism genes. In the vast majority of NF-Y-bound promoters, where NF-Y synergizes with neighboring TFs, it appears to be more of a promoter organizer and facilitator of transcription than a strong activator per se. Our results strongly suggest that cooperativity mediated by

precise spacing is a general mechanism utilized by NF-Y to regulate transcription of its target genes.

Conclusions

Our comprehensive analysis of NF-Y confirms many functions including its prevalence at proximal promoters, particularly those of growth controlling genes, at a much higher degree of precision and completion. More interestingly, our analyses uncover several novel and unexpected aspects of NF-Y function. In particular, NF-Y binds asymmetrically at its target sites, plays an important role at many tissue-specific enhancers, is capable of binding "closed" chromatin including at LTRs, coassociates pervasively with FOS but not other AP-1 factors, and displays precise stereo positioning with a restricted group of TFs involved in cellular proliferation. Lastly, we note that comprehensive bioinformatic analyses of the type performed here have been done on relatively few TFs. Similar analyses on other TFs whose target sites have been or will be defined by ChIP-seq are likely to uncover new functional properties and relationships of biological relevance.

Methods

Cell culture

K562, GM12878, and HeLa S3 were grown as per standard ENCODE protocols (The ENCODE Project Consortium 2011) and a detailed protocol is available at http://genome.ucsc.edu/ENCODE/.

ENCODE data sets

ChIP-sequencing data sets for histone PTMs, TFs, and RNA-seq for K562 and/or HeLa S3 cell lines were provided by ENCODE via the UCSC Genome Browser and are described there and elsewhere (The ENCODE Project Consortium 2011; http://genome.ucsc.edu/ENCODE/). ChIP-seq data sets were mapped and peaks called as described in the Supplemental Methods. RNA-seq data was prepared by Helicos as long (>200 nt), poly(A)-enriched, cytosolic RNA, and mapped using rSeq (Jiang and Wong 2008, 2009). Chromatin state maps and the associated numbering are from ENCODE and are detailed at http://genome.ucsc.edu/ENCODE/ and in Ernst et al. (2011). The chromatin state "heterochromatin" was renamed to "non-modified-chromatin."

Lentiviral knockdown and gene expression arrays

Scrambled control (shSCM) and NF-YA pLKO.1-shRNAs were designed by Sigma-Aldrich. The puromycin resistance cassette was replaced with an EGFP cassette. Viral production and transduction were carried out as previously described (Benatti et al. 2011). HeLa S3 cells were transduced with shSCM (scrambled control) or shNF-YA viral supernatants, in triplicate, and cells were collected after 48 h of incubation. The distribution of cells within the cell cycle was checked via FACS as previously described (Benatti et al. 2011). Knockdown efficiency was assayed by PCR on cDNA to known
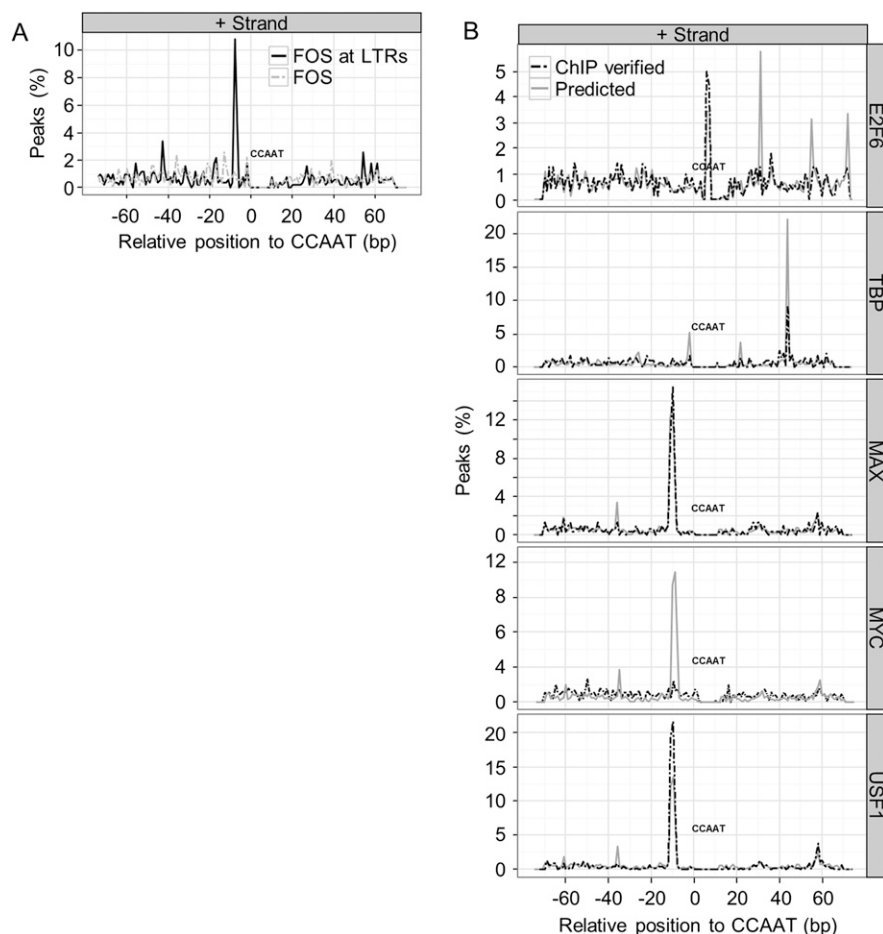
Figure 8. Motif pairings with the CCAAT motif are stereo positioned. (*A*) The percentage of NF-Y sites that have an AP-1 motif at the specified distance from the best scoring CCAAT motif centered at 0 bp. NF-YB peaks overlapping LTRs are categorized as "predicted," while the subset of NF-YB sites overlapping the respective ChIP-seq sites of FOS are categorized as "verified." The negative strand plots are near identical mirror images of the positive strand plots and are not shown. (*B*) Similar to *A*, except that all genomic regions are considered. The percentage of NF-YB peaks that have a TATA element (TBP), E box (MYC, MAX, USF1), and E2F motif (E2F6) are plotted. All NF-Y peaks are categorized as "predicted," while those NF-Y peaks overlapping the respective ChIP-seq peaks of the other TF are categorized as "verified." Only the top 500 peaks in each category are plotted.

NF-YA target genes and by Western Blot on whole-cell protein extracts using anti-NF-YA and anti-actin antibodies. For arrays, total RNA was prepared by TRIzol extraction and Qiagen RNeasy kit purification, converted to biotinylated aRNA, and hybridized to U133 Plus 2.0 GeneChip expression arrays using the 3′ IVT Express Kit (Affymetrix) following the manufacturer's protocol. Arrays were RMA normalized (Irizarry et al. 2003), gene expression levels calculated, differential expression determined, and probes annotated using the following R packages from the Bioconductor project: *affy* (Gautier et al. 2004), *limma* (Smyth 2004), and *annaffy* (http://www.bioconductor.org/packages/devel/bioc/html/annaffy.html).

Annotation of peaks to gene features, GO analysis (GREAT/IPA)

Genomic locations of peak summits (where summit is the local maxima in read counts) were submitted to the annotation tool GREAT (McLean et al. 2010) using the following parameters: whole-genome background set, basal plus extension, proximal upstream, 5 kbp; proximal downstream, 1 kbp; distal, 1 mbp; or whole-genome background set, basal, proximal upstream, 5 kbp;

proximal downstream, 1 kbp. Molecular signaling pathways were visualized using IPA (Ingenuity Systems: http://www.ingenuity.com) where a gray-shaded node represents a K562 NF-YB binding site located within the putative regulatory region, as defined by GREAT, of that molecule. Peak summits were annotated to genomic features using in-house scripts.

Motif stereo-positioning

NF-YB summit locations from K562 were scanned using Pscan (Zambelli et al. 2009) for matches to the NF-Y matrix in the JASPAR_CORE_2009 database (MA0060.1) (Portales-Casamar et al. 2010). For NF-Y loci with the best matrix match on the positive strand, the first C (of CCAAT) of the best match was set to 0 bp. Genomic sequences ±75 bp from the motifs were retrieved and scanned with Pscan using the collection of matrices in the JASPAR_CORE_2009 database (Portales-Casamar et al. 2010). For each JASPAR matrix, only regions containing a best matrix match >0.8 (computed as described in Zambelli et al. 2009) were considered for further analyses. This population was deemed "predicted." For each "predicted" population, the subpopulation of regions that overlapped the relevant TF ChIP-seq peak data set were deemed "ChIP verified." The frequency of the best motif occurrences for each motif matrix at each base pair from the CCAAT motif was determined for each population and plotted as the percentage of motifs.

Histone modifications and chromatin-associated factor clustering

Density arrays at NF-YB peak summits spanning either ±5 kbp or ±500 bp representing ChIP-seq read counts of histone PTMs (H3K79me2, H3K4me3, H3K27me3, H3K4me1, H4K20me1, H3K36me3, H3K4me2, H3K9ac, H3K9me1, H3K27ac), NF-YA, NF-YB, and RNA Pol II or NF-YA, NF-YB, and 78 chromatin-associated factors (see Supplemental Fig. 15A for the full list) with appropriate input samples, were computed using the ranked based correlation method of seqMINER v1.2 (Ye et al. 2011). Clustering was carried out using the following parameters: T = 10, K-means. Clusters from three to 50 were considered. Non-normalized raw read counts are depicted in Figure 3A and Supplemental Figures 7, 8, and 15A.

Mapping to repeats

Bowtie (Langmead et al. 2009) was used to map the NF-YB and input ChIP-seq data sets to a reference genome composed of Repbase v15.08 (Jurka et al. 2005) entries—simple.ref, humrep.ref, humsub.ref, and pseudo.ref—allowing ≤2 mismatches per read, and reads with >1 alignment had one alignment selected at random. Read counts for each Repbase entry were tallied and the ChIP:input ratio calculated. Individual consensus sequences of repeat elements were scored for the presence or absence of the

CCAAT motif using the matrix derived from this study and FIMO (Grant et al. 2011), with matches called at a significance P-value threshold of 10^{-4}.

Hierarchical clustering of binding events to promoters and enhancers

Regions considered promoters and enhancers were taken from the K562 chromatin state maps of Ernst et al. (2011). Regions were considered "bound" if an NF-YB peak summit directly overlapped the region. Regions were considered "nonbound" if no NF-YB peak overlapped the region of interest and the region had $<1.5\times$ the normalized fold-over-input ChIP-seq enrichment. At all NF-YB bound or NF-YB nonbound regions, chromatin associated factors were scored as present (1) or absent (0) based on directly overlapping peak summits. The R packages *pvclust* (Suzuki and Shimodaira 2006) and *snow* (http://cran.r-project.org/web/packages/snow/) were used to cluster the matrices and to calculate P-values using multiscale bootstrap resampling. Parameters were: method.dist="binary", method.hclust="ward", nboot=10000. Red and blue numbers in plots indicate the approximately unbiased (AU) P-values and the bootstrap probability (BP), respectively, as detailed in Suzuki and Shimodaira (2006).

Statistical test of TF coassociation with NF-YB

NF-YB-bound regions were as above. We assessed promoters or enhancers occupied by NF-YB for individual co-occupancy of 78 transcriptional regulators. The significance of the overlap was tested by a 2×2 contingency table using Fisher's exact test and calculated (Carlson et al. 2009), and P-values $<10^{-9}$ were deemed significant.

Data access

Microarray gene expression data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE40215.

Acknowledgments

References

Baxevanis AD, Arents G, Moudrianakis EN, Landsman D. 1995. A variety of DNA-binding and multimeric proteins contain the histone fold motif. *Nucleic Acids Res* **23:** 2685–2691.

Benachenhou F, Blikstad V, Blomberg J. 2009a. The phylogeny of orthoretroviral long terminal repeats (LTRs). *Gene* **448:** 134–138.

Benachenhou F, Jern P, Oja M, Sperber G, Blikstad V, Somervuo P, Kaski S, Blomberg J. 2009b. Evolutionary conservation of orthoretroviral long terminal repeats (LTRs) and ab initio detection of single LTRs in genomic data. *PLoS ONE* **4:** e5179.

Benatti P, Dolfini D, Vigano A, Ravo M, Weisz A, Imbriano C. 2011. Specific inhibition of NF-Y subunits triggers different cell proliferation defects. *Nucleic Acids Res* **39:** 5356–5368.

Bhattacharya A, Deng JM, Zhang Z, Behringer R, de Crombrugghe B, Maity SN. 2003. The B subunit of the CCAAT box binding transcription factor complex (CBF/NF-Y) is essential for early mouse development and cell proliferation. *Cancer Res* **63:** 8167–8172.

Bi W, Wu L, Coustry F, de Crombrugghe B, Maity SN. 1997. DNA binding specificity of the CCAAT-binding factor CBF/NF-Y. *J Biol Chem* **272:** 26562–26572.

Bourque G. 2009. Transposable elements in gene regulation and in the evolution of vertebrate genomes. *Curr Opin Genet Dev* **19:** 607–612.

Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew JL, Ruan Y, Wei CL, Ng HH, et al. 2008. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res* **18:** 1752–1762.

Caretti G, Motta MC, Mantovani R. 1999. NF-Y associates with H3-H4 tetramers and octamers by multiple mechanisms. *Mol Cell Biol* **19:** 8591–8603.

Carlson JM, Heckerman D, Shani G. 2009. Estimating false discovery rates for contingency tables. *Microsoft Research Technical Reports* MSR-TR-2009-53.

Ceribelli M, Alcalay M, Vigano MA, Mantovani R. 2006. Repression of new p53 targets revealed by ChIP on chip experiments. *Cell Cycle* **5:** 1102–1110.

Ceribelli M, Dolfini D, Merico D, Gatta R, Vigano AM, Pavesi G, Mantovani R. 2008. The histone-like NF-Y is a bifunctional transcription factor. *Mol Cell Biol* **28:** 2047–2058.

Dolfini D, Zambelli F, Pavesi G, Mantovani R. 2009. A perspective of promoter architecture from the CCAAT box. *Cell Cycle* **8:** 4127–4137.

Dolfini D, Gatta R, Mantovani R. 2012. NF-Y and the transcriptional activation of CCAAT promoters. *Crit Rev Biochem Mol Biol* **47:** 29–49.

Dorn A, Fehling HJ, Koch W, Le Meur M, Gerlinger P, Benoist C, Mathis D. 1988. B-cell control region at the 5′ end of a major histocompatibility complex class II gene: Sequences and factors. *Mol Cell Biol* **8:** 3975–3987.

Dutta A, Stoeckle MY, Hanafusa H. 1990. Serum and v-*src* increase the level of a CCAAT-binding factor required for transcription from a retroviral long terminal repeat. *Genes Dev* **4:** 243–254.

The ENCODE Project Consortium. 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306:** 636–640.

The ENCODE Project Consortium. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447:** 799–816.

The ENCODE Project Consortium. 2011. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* **9:** e1001046.

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

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, et al. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473:** 43–49.

Evans PM, Zhang W, Chen X, Yang J, Bhakat KK, Liu C. 2007. Kruppel-like factor 4 is acetylated by p300 and regulates gene transcription via modulation of histone acetylation. *J Biol Chem* **282:** 33994–34002.

Faber M, Sealy L. 1990. Rous sarcoma virus enhancer factor I is a ubiquitous CCAAT transcription factor highly related to CBF and NF-Y. *J Biol Chem* **265:** 22243–22254.

Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A, et al. 2011. The UCSC Genome Browser database: Update 2011. *Nucleic Acids Res* **39:** D876–D882.

Gatta R, Mantovani R. 2011. NF-Y affects histone acetylation and H2A.Z deposition in cell cycle promoters. *Epigenetics* **6:** 526–534.

Gautier L, Cope L, Bolstad BM, Irizarry RA. 2004. affy–analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20:** 307–315.

Gilthorpe J, Vandromme M, Brend T, Gutman A, Summerbell D, Totty N, Rigby PW. 2002. Spatially specific expression of Hoxb4 is dependent on the ubiquitous transcription factor NFY. *Development* **129:** 3887–3899.

Goodarzi H, Elemento O, Tavazoie S. 2009. Revealing global regulatory perturbations across human cancers. *Mol Cell* **36:** 900–911.

Grant CE, Bailey TL, Noble WS. 2011. FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **27:** 1017–1018.

Graves BJ, Johnson PF, McKnight SL. 1986. Homologous recognition of a promoter domain common to the MSV LTR and the HSV tk gene. *Cell* **44:** 565–576.

Greuel BT, Sealy L, Majors JE. 1990. Transcriptional activity of the Rous sarcoma virus long terminal repeat correlates with binding of a factor to an upstream CCAAT box in vitro. *Virology* **177:** 33–43.

Gubler U, Chua AO, Schoenhaut DS, Dwyer CM, McComas W, Motyka R, Nabavi N, Wolitzky AG, Quinn PM, Familletti PC, et al. 1991. Coexpression of two distinct genes is required to generate secreted bioactive cytotoxic lymphocyte maturation factor. *Proc Natl Acad Sci* **88:** 4143–4147.

Gurtner A, Fuschi P, Martelli F, Manni I, Artuso S, Simonte G, Ambrosino V, Antonini A, Folgiero V, Falcioni R, et al. 2010. Transcription factor NF-Y induces apoptosis in cells expressing wild-type p53 through E2F1 upregulation and p53 activation. *Cancer Res* **70:** 9711–9720.

He S, Liu C, Skogerbo G, Zhao H, Wang J, Liu T, Bai B, Zhao Y, Chen R. 2008. NONCODE v2.0: Decoding the non-coding. *Nucleic Acids Res* **36:** D170–D172.

Hooft van Huijsduijnen RA, Bollekens J, Dorn A, Benoist C, Mathis D. 1987. Properties of a CCAAT box-binding protein. *Nucleic Acids Res* **15:** 7265–7282.

Huang S, Li X, Yusufzai TM, Qiu Y, Felsenfeld G. 2007. USF1 recruits histone modification complexes and is critical for maintenance of a chromatin barrier. *Mol Cell Biol* **27:** 7991–8002.

Hughes R, Kristiansen M, Lassot I, Desagher S, Mantovani R, Ham J. 2011. NF-Y is essential for expression of the proapoptotic bim gene in sympathetic neurons. *Cell Death Differ* **18:** 937–947.

Imbriano C, Gnesutta N, Mantovani R. 2012. The NF-Y/p53 liaison: Well beyond repression. *Biochim Biophys Acta* **1825:**131–139.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4:** 249–264.

Izumi H, Molander C, Penn LZ, Ishisaki A, Kohno K, Funa K. 2001. Mechanism for the transcriptional repression by c-Myc on PDGF β-receptor. *J Cell Sci* **114:** 1533–1544.

Jiang H, Wong WH. 2008. SeqMap: Mapping massive amount of oligonucleotides to the genome. *Bioinformatics* **24:** 2395–2396.

Jiang H, Wong WH. 2009. Statistical inferences for isoform expression in RNA-Seq. *Bioinformatics* **25:** 1026–1032.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* **110:** 462–467.

Kalra IS, Alam MM, Choudhary PK, Pace BS. 2011. Kruppel-like Factor 4 activates HBG gene expression in primary erythroid cells. *Br J Haematol* **154:** 248–259.

Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, et al. 2009. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci* **106:** 11667–11672.

Kim CG, Swendeman SL, Barnhart KM, Sheffery M. 1990. Promoter elements and erythroid cell nuclear factors that regulate α-globin gene transcription in vitro. *Mol Cell Biol* **10:** 5958–5966.

Kim IS, Sinha S, de Crombrugghe B, Maity SN. 1996. Determination of functional domains in the C subunit of the CCAAT-binding factor (CBF) necessary for formation of a CBF-DNA complex: CBF-B interacts simultaneously with both the CBF-A and CBF-C subunits to form a heterotrimeric CBF molecule. *Mol Cell Biol* **16:** 4003–4013.

Kozomara A, Griffiths-Jones S. 2011. miRBase: Integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* **39:** D152–D157.

Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G. 2010. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet* **42:** 631–634.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10:** R25.

Li X, Wang S, Li Y, Deng C, Steiner LA, Xiao H, Wu C, Bungert J, Gallagher PG, Felsenfeld G, et al. 2011. Chromatin boundaries require functional collaboration between the hSET1 and NURF complexes. *Blood* **118:** 1386–1394.

Litovchick L, Sadasivam S, Florens L, Zhu X, Swanson SK, Velmurugan S, Chen R, Washburn MP, Liu XS, DeCaprio JA. 2007. Evolutionarily conserved multisubunit RBL2/p130 and E2F4 protein complex represses human cell cycle-dependent genes in quiescence. *Mol Cell* **26:** 539–551.

Magnani L, Eeckhoute J, Lupien M. 2011. Pioneer factors: Directing transcriptional regulators within the chromatin environment. *Trends Genet* **27:** 465–474.

Maksakova IA, Mager DL, Reiss D. 2008. Keeping active endogenous retroviral-like elements in check: The epigenetic perspective. *Cell Mol Life Sci* **65:** 3329–3347.

Marino-Ramirez L, Spouge JL, Kanga GC, Landsman D. 2004. Statistical analysis of over-represented words in human promoter sequences. *Nucleic Acids Res* **32:** 949–958.

Martynova E, Pozzi S, Basile V, Dolfini D, Zambelli F, Imbriano C, Pavesi G, Mantovani R. 2012. Gain-of-function p53 mutants have widespread genomic locations partially overlapping with p63. *Oncotarget* **3:** 132–143.

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. 2010. GREAT improves functional interpretation of *cis*-regulatory regions. *Nat Biotechnol* **28:** 495–501.

Moqtaderi Z, Wang J, Raha D, White RJ, Snyder M, Weng Z, Struhl K. 2010. Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells. *Nat Struct Mol Biol* **17:** 635–640.

Morachis JM, Murawsky CM, Emerson BM. 2010. Regulation of the p53 transcriptional response by structurally diverse core promoters. *Genes Dev* **24:** 135–147.

Muller GA, Engeland K. 2010. The central role of CDE/CHR promoter elements in the regulation of cell cycle-dependent gene transcription. *FEBS J* **277:** 877–893.

Muller GA, Quaas M, Schumann M, Krause E, Padi M, Fischer M, Litovchick L, DeCaprio JA, Engeland K. 2012. The CHR promoter element controls cell cycle-dependent gene transcription and binds the DREAM and MMB complexes. *Nucleic Acids Res* **40:** 1561–1578.

Nardini M, Gnesutta N, Donati G, Gatta R, Forni C, Fossati A, Vonrhein C, Moras D, Romier C, Bolognesi M, et al. 2013. Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. *Cell* **152:** 132–143.

Oishi Y, Manabe I, Tobe K, Ohsugi M, Kubota T, Fujiu K, Maemura K, Kubota N, Kadowaki T, Nagai R. 2008. SUMOylation of Kruppel-like transcription factor 5 acts as a molecular switch in transcriptional programs of lipid metabolism involving PPAR-δ. *Nat Med* **14:** 656–666.

Pi W, Zhu X, Wu M, Wang Y, Fulzele S, Eroglu A, Ling J, Tuan D. 2010. Long-range function of an intergenic retrotransposon. *Proc Natl Acad Sci* **107:** 12992–12997.

Portales-Casamar E, Thongjuea S, Kwon AT, Arenillas D, Zhao X, Valen E, Yusuf D, Lenhard B, Wasserman WW, Sandelin A. 2010. JASPAR 2010: The greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res* **38:** D105–D110.

Ravasi T, Suzuki H, Cannistraci CV, Katayama S, Bajic VB, Tan K, Akalin A, Schmeier S, Kanamori-Katayama M, Bertin N, et al. 2010. An atlas of combinatorial transcriptional regulation in mouse and man. *Cell* **140:** 744–752.

Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Barrette TR, Ghosh D, Chinnaiyan AM. 2005. Mining for regulatory programs in the cancer transcriptome. *Nat Genet* **37:** 579–583.

Romier C, Cocchiarella F, Mantovani R, Moras D. 2003. The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y. *J Biol Chem* **278:** 1336–1345.

Salsi V, Caretti G, Wasner M, Reinhard W, Haugwitz U, Engeland K, Mantovani R. 2003. Interactions between p300 and multiple NF-Y trimers govern cyclin B2 promoter function. *J Biol Chem* **278:** 6642–6650.

Scheef G, Fischer N, Flory E, Schmitt I, Tonjes RR. 2002. Transcriptional regulation of porcine endogenous retroviruses released from porcine and infected human cells by heterotrimeric protein complex NF-Y and impact of immunosuppressive drugs. *J Virol* **76:** 12553–12563.

Schmit F, Korenjak M, Mannefeld M, Schmitt K, Franke C, von Eyss B, Gagrica S, Hanel F, Brehm A, Gaubatz S. 2007. LINC, a human complex that is related to pRB-containing complexes in invertebrates regulates the expression of G₂/M genes. *Cell Cycle* **6:** 1903–1913.

Schneider TD, Stephens RM. 1990. Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res* **18:** 6097–6100.

Schuierer M, Hilger-Eversheim K, Dobner T, Bosserhoff AK, Moser M, Turner J, Crossley M, Buettner R. 2001. Induction of *AP-2α* expression by adenoviral infection involves inactivation of the AP-2rep transcriptional corepressor CtBP1. *J Biol Chem* **276:** 27944–27949.

Sinha S, Maity SN, Lu J, de Crombrugghe B. 1995. Recombinant rat CBF-C, the third subunit of CBF/NFY, allows formation of a protein-DNA complex with CBF-A and CBF-B and with yeast HAP2 and HAP3. *Proc Natl Acad Sci* **92:** 1624–1628.

Sinha S, Kim IS, Sohn KY, de Crombrugghe B, Maity SN. 1996. Three classes of mutations in the A subunit of the CCAAT-binding factor CBF delineate functional domains involved in the three-step assembly of the CBF-DNA complex. *Mol Cell Biol* **16:** 328–337.

Sinha S, Adler AS, Field Y, Chang HY, Segal E. 2008. Systematic functional characterization of *cis*-regulatory motifs in human core promoters. *Genome Res* **18:** 477–488.

Smyth GK. 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet MolBiol* **3**. doi: 10.2202/1544-6115.0127.

Strub T, Giuliano S, Ye T, Bonet C, Keime C, Kobi D, Le Gras S, Cormont M, Ballotti R, Bertolotto C, et al. 2011. Essential role of microphthalmia transcription factor for DNA replication, mitosis and genomic stability in melanoma. *Oncogene* **30:** 2319–2332.

Suzuki R, Shimodaira H. 2006. Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22:** 1540–1542.

Suzuki Y, Tsunoda T, Sese J, Taira H, Mizushima-Sugano J, Hata H, Ota T, Isogai T, Tanaka T, Nakamura Y, et al. 2001. Identification and

characterization of the potential promoter regions of 1031 kinds of human genes. *Genome Res* **11:** 677–684.

Tabach Y, Milyavsky M, Shats I, Brosh R, Zuk O, Yitzhaky A, Mantovani R, Domany E, Rotter V, Pilpel Y. 2005. The promoters of human cell cycle genes integrate signals from two tumor suppressive pathways during cellular transformation. *Mol Syst Biol* **1:** 2005.0022.

Testa A, Donati G, Yan P, Romani F, Huang TH, Vigano MA, Mantovani R. 2005. Chromatin immunoprecipitation (ChIP) on chip experiments uncover a widespread distribution of NF-Y binding CCAAT sites outside of core promoters. *J Biol Chem* **280:** 13606–13615.

Tiwari VK, Stadler MB, Wirbelauer C, Paro R, Schubeler D, Beisel C. 2012. A chromatin-modifying function of JNK during stem cell differentiation. *Nat Genet* **44:** 94–100.

Turner J, Crossley M. 1998. Cloning and characterization of mCtBP2, a co-repressor that associates with basic Krüppel-like factor and other mammalian transcriptional regulators. *EMBO J* **17:** 5129–5140.

van Vliet J, Turner J, Crossley M. 2000. Human Krüppel-like factor 8: A CACCC-box binding protein that associates with CtBP and represses transcription. *Nucleic Acids Res* **28:** 1955–1962.

West AG, Huang S, Gaszner M, Litt MD, Felsenfeld G. 2004. Recruitment of histone modifications by USF proteins at a vertebrate barrier element. *Mol Cell* **16:** 453–463.

Wolf SF, Temple PA, Kobayashi M, Young D, Dicig M, Lowe L, Dzialo R, Fitz L, Ferenz C, Hewick RM, et al. 1991. Cloning of cDNA for natural killer cell stimulatory factor, a heterodimeric cytokine with multiple biologic effects on T and natural killer cells. *J Immunol* **146:** 3074–3081.

Yang A, Zhu Z, Kapranov P, McKeon F, Church GM, Gingeras TR, Struhl K. 2006. Relationships between p63 binding, DNA sequence, transcription activity, and biological function in human cells. *Mol Cell* **24:** 593–602.

Ye T, Krebs AR, Choukrallah MA, Keime C, Plewniak F, Davidson I, Tora L. 2011. seqMINER: An integrated ChIP-seq data interpretation platform. *Nucleic Acids Res* **39:** e35.

Yoon HS, Yang VW. 2004. Requirement of Krüppel-like factor 4 in preventing entry into mitosis following DNA damage. *J Biol Chem* **279:** 5035–5041.

Yoshida H, Okada T, Haze K, Yanagi H, Yura T, Negishi M, Mori K. 2000. ATF6 activated by proteolysis binds in the presence of NF-Y (CBF) directly to the *cis*-acting element responsible for the mammalian unfolded protein response. *Mol Cell Biol* **20:** 6755–6767.

Yu X, Zhu X, Pi W, Ling J, Ko L, Takeda Y, Tuan D. 2005. The long terminal repeat (LTR) of ERV-9 human endogenous retrovirus binds to NF-Y in the assembly of an active LTR enhancer complex NF-Y/MZF1/GATA-2. *J Biol Chem* **280:** 35184–35194.

Zambelli F, Pesole G, Pavesi G. 2009. Pscan: Finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. *Nucleic Acids Res* **37:** W247–W252.

Zaret KS, Carroll JS. 2011. Pioneer transcription factors: Establishing competence for gene expression. *Genes Dev* **25:** 2227–2241.

Zemzoumi K, Frontini M, Bellorini M, Mantovani R. 1999. NF-Y histone fold α1 helices help impart CCAAT specificity. *J Mol Biol* **286:** 327–337.