SUPPLEMENTAL METHODS

ChIP. Cells were fixed by the addition of 1% v/v formaldehyde at room temperature for 10 min and quenched with 0.2 M glycine. Cell pellets were washed twice with PBS, lysed in CLB (25 mM HEPES pH7.8, 1.5 mM MgCl2, 10 mM KCl, 0.1% NP-40) with 1 mM DTT added just before use, and nuclei pelleted by centrifugation at 12 kG. Crude nuclei were then lysed in NLB (50 mM HEPES pH7.9, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na deoxycholate) with 1% SDS. Nuclear extract was fragmented using a Branson 450 sonicator and/or Misonix 3000 to yield chromatin of a suitable length for immuno-precipitation. Chromatin was spun at 12 kG for 10 min to remove precipitates and the supernatant was flash frozen and stored at -80ºC until use. Chromatin from $2x10^7$ cell equivalents were used per ChIP. Chromatin was diluted 10x in NLB and pre-cleared with Protein A-Sepharose beads for 2 hr at 4ºC. The supernatant was incubated with 5-10 μg of the appropriate antibody overnight at 4ºC. Protein A-Sepharose beads were added for 2 hr then washed as follows: 2x NLB with 0.1% SDS; 2x NLB with 0.1% SDS and 640 mM NaCl; 2x WB (20 mM Tris-HCL pH8.0, 250 mM LiCl, 1 mM EDTA pH8.0, 0.5% NP-40, 0.5% Na deoxycholate); and, 2x TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH8.0). Bound protein was eluted in TE with 1% SDS for 15 min at >60ºC. Protease inhibitor cocktail and 1 mM PMSF were added to all buffers just before use. Elutions were digested with 20 μL of 20 mg/mL pronase for 2 hr at 42ºC and crosslinks reversed by overnight incubation at 65ºC. DNA was purified using phenol:chloroform:isoamyl alcohol extraction utilizing high density MaXtract tubes (Qiagen, USA) as per manufacturers protocol. Aqueous phase DNA was precipitated by the addition of 200 mM NaCl, 500 mM NaAc, 80 μg/mL glycogen, and 2

volumes of ethanol, while incubating for >1 hr at -80°C, followed by centrifugation at 12 kG at 4°C for 20 min. The precipitate was washed in 95% ethanol, resuspended in TE and stored at -20°C until needed.

ChIP-Sequencing. ChIP DNA (2 biological replicates) prepared as above, and immuno-precipitated anti-NF-YB or anti-NF-YA antibody (Mantovani, R.), and input DNA (3 biological replicates) were end repaired with calf intestinal alkaline phosphatase (New England Biolabs, USA) and sent for sequencing to the Stanford Center for Genomics and Personalized Medicine or the Department of Molecular, Cellular and Developmental Biology at Yale University. Library preparation and Illumina sequencing were carried out as per manufacturer protocols and ENCODE standards ((The ENCODE Project Consortium 2011), http://genome.ucsc.edu/ENCODE/). Sequence reads (28 nucleotides) were mapped to the human genome (hg18) using *Bowtie* (Langmead et al. 2009), allowing <= 2 mismatches per read and reads with >10 reportable alignments were discarded. Binding sites were called using *MACS* v1.4 (Zhang et al. 2008) at a *p*-value threshold of 10^{-9} (unless otherwise noted) on non-redundant reads using input to control for local genomic biases.

ChIP-qPCR. Primer pairs were designed to amplify regions within 150 bp of the summit of ChIP-Seq peaks. *Batch primer3* was used for primer design using default parameters (You et al. 2008). All primers were tested for unique hits to the human genome using UCSC *In-Silico PCR* (Jim Kent, UCSC) and by dissociation curve analysis. See Supplemental Data for primer sequences. qPCR was performed on an Applied Biosystems 7900FAST instrument (kindly provided by the HMS ICCB) on ChIP and input DNA (prepared as above, except

Qiagen columns were used for purification), using 2x Taq Mix (Aparicio et al. 2005), except 250 nM EVA green (Biotium, USA) replaced SYBR green. PCR program was: 95°C 10 min, followed by 40 cycles of 95°C for 5 sec, 60°C for 30 sec. ChIP-qPCR values are represented as fold enrichment over an NF-Y non-bound control region as previously described (Cawley et al. 2004). Error bars are based on the standard deviation observed in 2-4 biological replicates run in qPCR triplicates.

Western blot and RT-PCR. As described in (Benatti et al. 2011). Briefly, total cell protein extracts were prepared by resuspending the cell pellets from siSCM or siNF-YA infected cells in lysis buffer (50 mM Tris–HCl pH 8.0, 120 mM NaCl, 0.5% NP-40, 1 mM EDTA, protease and phosphatase inhibitors). An equivalent amount of cellular extracts were resolved by SDS–PAGE, electro transferred to PVDF membrane, and immuno-blotted with the following antibodies at 1:1000 in TBS containing 1 mg/ml BSA: anti-NF-YA (sc-10779), and anti-actin (sc-1616) from Santa Cruz Biotechnology, USA.

De novo **motif discovery.** DNA sequences corresponding to the regions +/-50 bp of ChIP-Seq peak summits for NF-YB K562 were gathered using *BEDTools* (Quinlan and Hall 2010) and repeat masked using *RepeatMasker* with –q option (Smit et al. 1996-2010). Sequences were searched for *de novo* motifs using parallel *MEME* (Bailey and Elkan 1994) using the following parameters: -zoops, -revcomp, -minw (range 5-40), -maxw (range 10-60). Background letter frequencies were based on a 5-order Markov model derived from hg18 repeat masked sequences -350:+100 bp about RefSeq TSSs, or the non-modified-chromatin state (Ernst et al. 2011), as appropriate. For NF-YB, a second background model using

FAIRE-Seq regions from K562 was also carried out and produced a similar motif (not shown). *Tomtom* (Gupta et al. 2007) was used to compare *de novo* motifs to known motifs in the JASPAR_CORE_2009 database (Portales-Casamar et al. 2010). For TFs other than NF-Y, motifs were discovered as above except the top 1000 ChIP-Seq peaks of each factor were used and the top motif was selected, except for FOS, which produced the CCAAT box and was substituted for the motif derived from JUN ChIP-Seq. Similarly, for motifs in the non-modified-chromatin state *MEME* was run using a motif width range of 10-15, on all K562 NF-YB peaks residing within the non-modified-chromatin state, on non-masked sequences, with a background set derived from the entire non-modified-chromatin state of K562 (Ernst et al. 2011).

Data Access

Microarray gene expression datasets are deposited in GEO (GSE40215).

Supplemental Figure 1: NF-Y ChIP-Seq

A. Western blots of nuclear extracts from five cell lines probed with anti-NF-YA and anti-NF-YB antibodies. Arrows highlight doublet bands showing that both isoforms of NF-YA and NF-YB are detected.

B. Immuno-precipitation Western blot (IP-WB) of nuclear extracts showing enrichment of NF-YA and NF-YB specific bands in the elution and depletion in the supernatant. An IgG antibody was used as control for non-specific binding.

C. ChIP-qPCR results from anti-NF-YA and anti-NF-YB IPs performed in K562, GM12878 and HeLaS3, before sequencing, showing enrichment over an NF-Y non-bound control region.

D. Representative loci showing NF-YA, NF-YB and input control ChIP-Seq data from K562, GM12878 and HeLaS3. Enrichment of reads at the HNRNPA1 and SON promoters are specific to NF-Y and not present in the input dataset. GAPDH up. and TLE6 up. are control regions not bound by NF-Y. Red bars indicate ChIP-qPCR primer locations. Blue bars under peaks show MACS called peak regions at the 10^{-9} p-value. RefSeq genes are illustrated.

Supplemental Figure 2: CCAAT motif frequency and NF-YB ChIP-Seq saturation analysis

A. Mean percentage of peaks with CCAAT motifs in K562 NF-YB peaks called at specific p-values. CCAAT motifs were called using *FIMO* at a p-value threshold of 10^{-4}. Similarly sized random genomic regions have a CCAAT motif rate of 8%.

B. The percentage of peaks, from the K562 NF-YB 10^{-5} peak list, that are successfully identified based on a random subsampling of ChIP-Seq reads.

Supplemental Figure 3: ChIP-qPCR validation of NF-YB binding in the absence of NF-YA

A. NF-YB peaks with a high NF-YB:NF-YA read count ratio were checked by ChIP-qPCR. Selection criteria were: did not overlap an NF-YA peak called at a lenient *p*-value threshold of 10^{-5}; and, hand-checked by observation of raw ChIP-Seq data and discarded if appreciable NF-YA signal was present. A group of control targets that showed similar fold enrichments for NF-YB as the test group were selected for comparison. The ratio of NF-YB:NF-YA reads is shown, and targets are sorted by ratio. The average of 2-4 biological replicates and their associated standard deviations are depicted.

B. Distribution of normalized ratios of NF-YA and NF-YB ChIP-Seq read counts at NF-YB peaks. Reads were counted within a region spanning +/-100 bp from the summit of NF-YB peaks and normalized to the total number of mapped reads.

Supplemental Figure 4: NF-Y binds to genes involved in cell signaling, DNA repair, cell-cycle, and gene expression

A and B. Gene ontology analyses of all NF-YB sites in K562, GM12878, HeLaS3 (A), and sites common to and unique to individual cell types (B). Only the top 10 most significant terms with a fold enrichment > 2 are shown. Observed region hits corresponds to the number of regulatory regions of genes in that gene ontology term that had >= 1 NF-YB sites. Highly

redundant categories are not shown. For a full list of significant GO terms for each cell line see Supplemental Data Table 10, 11, and 12. Venn diagram is derived from Fig. 1D.

Supplemental Figure 5: NF-Y binds to many genes involved in transcriptional regulation and TP53 and TRAIL signaling

A and B. Transcriptional regulatory complexes, TFs, Pol II general factors, and chromatin associated factors and complexes whose members' putative *cis*-regulatory domains are bound by NF-YB. Dark and light green shading indicate NF-YB binding within -5 kbp: +1 kbp of TSS and -5 kbp: +1 kbp of TSS plus up to +/-1 mbp extension, respectively. Genes and categories are derived from transcription related GO terms that are significantly targeted by NF-Y. For a full list of significant GO terms, see Supplemental Data table 10, 11, 12.

C and D. Ingenuity pathway analysis, showing the TP53 (C) and death (TRAIL) receptor (D) signaling pathways. Gray shaded gene terms indicate that that gene's putative *cis*-regulatory domain (-5 kbp: +1 kbp of TSS plus up to +/-1 mbp extension) was bound by NF-YB in K562.

Supplemental Figure 6: Annotation of NF-Y ChIP-Seq peaks to RefSeq gene features

Percent occurrence of K562 NF-YB peaks (black) at RefSeq gene features compared to the frequency of features in the human genome (gray). *P*-values are indicated.

Supplemental Figure 7: MYC bound loci do not reside within non-modified chromatin domains

K-means clustering of K562 MYC sites based on the distribution of histone PTMs, Pol II, NF-YA, and NF-YB ChIP-Seq reads within a region spanning +/- 5kbp from the summit of MYC peaks (centered at 0 bp). Clustering was carried out on transformed normalized read counts. Raw read count intensity is depicted in red. Similar to Fig. 3A.

Supplemental Figure 8: HeLaS3 NF-Y bound loci reside within 5 disparate epigenetic domains

K-means clustering of HeLaS3 NF-YB loci based on the distribution of histone PTMs, 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. Similar to Fig. 3A.

Supplemental Figure 9: NF-Y cell line specific sites are enriched for enhancers and function in cell type specific biological processes

A. Ratio of enhancer: promoter chromatin states in the GM12878 and K562 cell type specific NF-YB binding sites and sites common to all three cell types (K562, GM12878 and HeLaS3). Peaks are considered unique to a cell line if they do not overlap a peak, called at the lenient 10^{-5} p-value threshold, in the other two cell lines.

B. Box plot showing the distance to the nearest RefSeq TSS of NF-YB sites. Horizontal edges of the box represent the inter-quartile range. The middle bar represents the median value. Ends of the extensions represent the minimum and maximum datum within 1.5 x inter-quartile range. Outliers are represented as dots. p-values represent the significance of the difference in the median value calculated by the Wilcoxon rank sum test.

Supplemental Figure 10: Functional inactivation of NF-YA by shRNA knockdown

Representative semi-quantitative PCR (A) and Western blot (B) analysis of an NF-YA and scrambled control lentiviral shRNA knockdown in HeLaS3.

A. *GAPDH* is not known to be regulated by NF-Y (negative control), whereas *CCNB1* and *TOPOIIA* are known NF-Y regulated genes (positive controls). A reverse transcriptase negative control (RT-) is also shown.

B. Membranes were blotted with anti-NF-YA or anti-actin antibodies (control) and show a specific reduction in NF-YA protein levels.

C. Differentially expressed genes upon NF-YA knockdown in HeLaS3 and the number that are bound by NF-Y as determined by ChIP-Seq. Distance windows are in relation to RefSeq TSSs. Adjusted *p*-value is Bonferroni corrected.

Supplemental Figure 11: Functional inactivation of NF-YA and correlation with ChIP-Seq NF-Y sites

A. The fraction (mean of 500 bp sliding window) of RefSeq TSSs with an NF-YA or NF-YB ChIP-Seq peak within the indicated range of the TSS, ranked according to differential gene expression upon shNF-YA of the associated genes.

B. NF-YA, NF-YB, or MYC sites, excluding peaks that overlapped LTRs, were mapped to the nearest RefSeq TSS and the distance and associated differential gene expression upon shNF-YA of that gene determined. Peaks are ranked based on differential gene expression and the median distance of a sliding 200 peak window is shown. Red arrows indicate the

presence of NF-Y sites that are greatly distant to the most down-regulated genes, which are not present when MYC sites are used.

Supplemental Figure 12: NF-Y binds to specific classes of LTRs

Number of NF-Y sites unique to K562, GM12878, and HeLaS3 or sites common to all three cell types at long terminal repeats.

Supplemental Figure 13: TFs have marked differences in their ability to bind their motif in closed chromatin

The percentage of genome-wide, computationally discovered, TF binding site motif locations within non-modified-chromatin, PcG repressed and strong promoter chromatin states, FAIRE-Seq regions or the entire genome, that directly overlap their ChIP-Seq TF peaks plotted as a function of motif quality (right axes). Similar to Fig. 5C.

Supplemental Figure 14: NF-Y significantly co-associates with many factors at promoters and enhancers

A. The number of promoters and enhancers bound by NF-YB and each individual chromatin associated protein was assayed by a 2x2 contingency table approach. The significance of the observed overlap was determined by the Fisher exact test. Peak summits from the 10^{-9} peak lists were used to determine occupancy within a given region.

B. Dendrograms depicting the correlation between chromatin associated factors at NF-YB bound and NF-YB non-bound promoters and enhancers in K562. All promoters and enhancers from the chromatin state maps where scored for the presence/absence of all

chromatin-associated factors based on the location of peak summits and clustered. NF-YA and NF-YB are indicated by arrows and the cluster they associate with is shaded in yellow.

C. Multi-way overlaps between chromatin associated factors (Pol II/III and associated general factors were not considered) at NF-YB bound and NF-YB non-bound strong promoters and all enhancers. Only the top 10 combinations are shown. The number and percentage of promoters or enhancers that are simultaneously bound by the indicated factor(s) are shown. Yellow shading represents FOS, which is highly prevalent at NF-YB bound promoters.

Supplemental Figure 15: NF-Y partners with FOS, USF1, USF2 and SP1 in non-modified-chromatin domains

A. K-means clustering of K562 NF-YB loci from the non-modified-chromatin class (clusters D and J; Fig. 3A) based on the distribution of ChIP-Seq reads from chromatin-associated factors within a region spanning +/-500 bp 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.

B. *De novo* motif search of NF-YB peaks in the non-modified-chromatin state. Only the top 5 motifs are shown. The respective best match (*p*-value shown to right) to known motifs are shown on top of the discovered motifs. The percentage of NF-YB loci containing the discovered motif is indicated to the right. The very similar HAP3 (yeast NF-Y orthologue) motif was replaced by the NF-YA motif which was second to HAP3 in all cases.

REFERENCES

Aparicio O, Geisberg JV, Sekinger E, Yang A, Moqtaderi Z, Struhl K. 2005. Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences in vivo. *Current protocols in molecular biology / edited by Frederick M Ausubel [et al]* **Chapter 21**: Unit 21 23.

Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings / International Conference on Intelligent Systems for Molecular Biology ; ISMB International Conference on Intelligent Systems for Molecular Biology* **2**: 28-36.

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 research* **39**(13): 5356-5368.

Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ et al. 2004. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116**(4): 499-509.

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**(7345): 43-49.

Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. 2007. Quantifying similarity between motifs. *Genome biology* **8**(2): R24.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**(3): R25.

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 research* **38**(Database issue): D105-110.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**(6): 841-842.

Smit AFA, Hubley R, Green P. 1996-2010. RepeatMasker Open-3.0.

The ENCODE Project Consortium. 2011. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS biology* **9**(4): e1001046.

You FM, Huo N, Gu YQ, Luo MC, Ma Y, Hane D, Lazo GR, Dvorak J, Anderson OD. 2008. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC bioinformatics* **9**: 253.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**(9): R137.