The American Journal of Human Genetics, Volume *100*

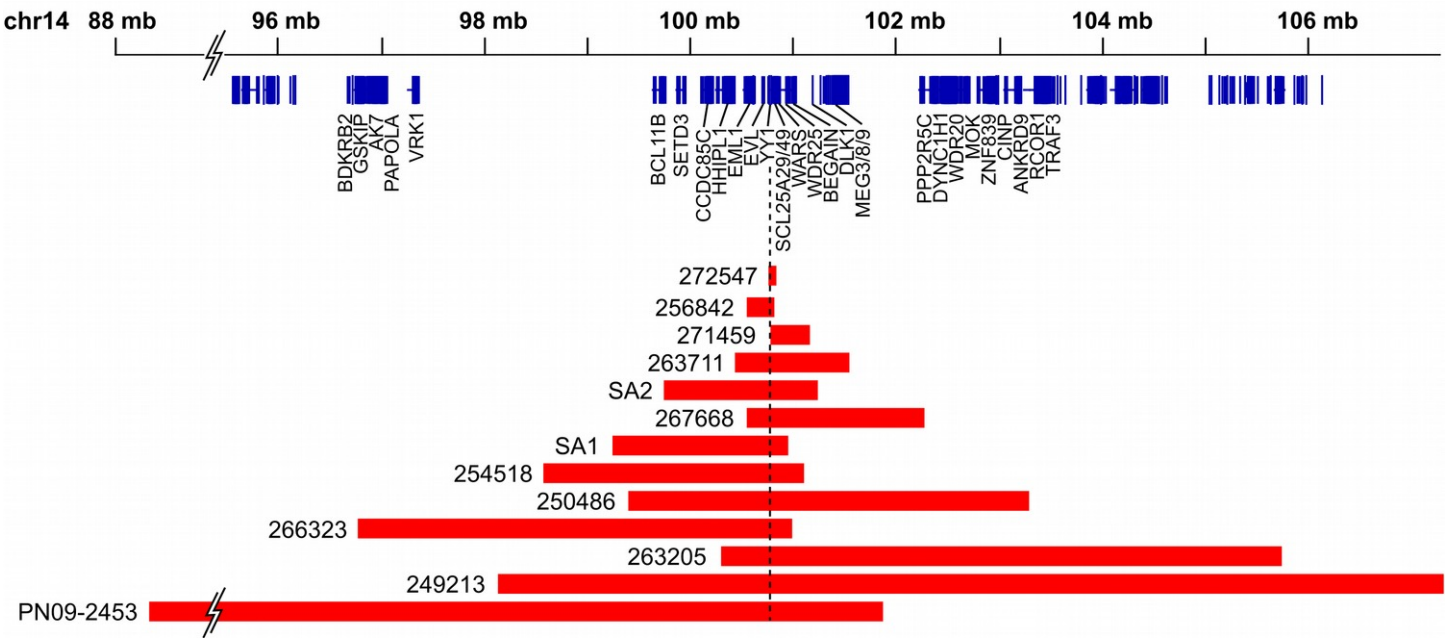
Supplemental Data

YY1 Haploinsufficiency Causes an Intellectual Disability Syndrome Featuring Transcriptional and Chromatin Dysfunction

Michele Gabriele, Anneke T. Vulto-van Silfhout, Pierre-Luc Germain, Alessandro Vitriolo, Raman Kumar, Evelyn Douglas, Eric Haan, Kenjiro Kosaki, Toshiki Takenouchi, Anita Rauch, Katharina Steindl, Eirik Frengen, Doriana Misceo, Christeen Ramane J. Pedurupillay, Petter Stromme, Jill A. Rosenfeld, Yunru Shao, William J. Craigen, Christian P. Schaaf, David Rodriguez-Buritica, Laura Farach, Jennifer Friedman, Perla Thulin, Scott D. McLean, Kimberly M. Nugent, Jenny Morton, Jillian Nicholl, Joris Andrieux, Asbjørg Stray-Pedersen, Pascal Chambon, Sophie Patrier, Sally A. Lynch, Susanne Kjaergaard, Pernille M. Tørring, Charlotte Brasch-Andersen, Anne Ronan, Arie van Haeringen, Peter J. Anderson, Zöe Powis, Han G. Brunner, Rolph Pfundt, Janneke H.M. Schuurs-Hoeijmakers, Bregje W.M. van Bon, Stefan Lelieveld, Christian Gilissen, Willy M. Nillesen, Lisenka E.L.M. Vissers, Jozef Gecz, David A. Koolen, Giuseppe Testa, and Bert B.A. de Vries

Figure S1

A



B

Figure S1: Depiction of the cases of deletions encompassing YY1. A: Representation of the genomic coordinates of the deletions. **B:** Frontal photographs of patients with deletions.

Figure S2

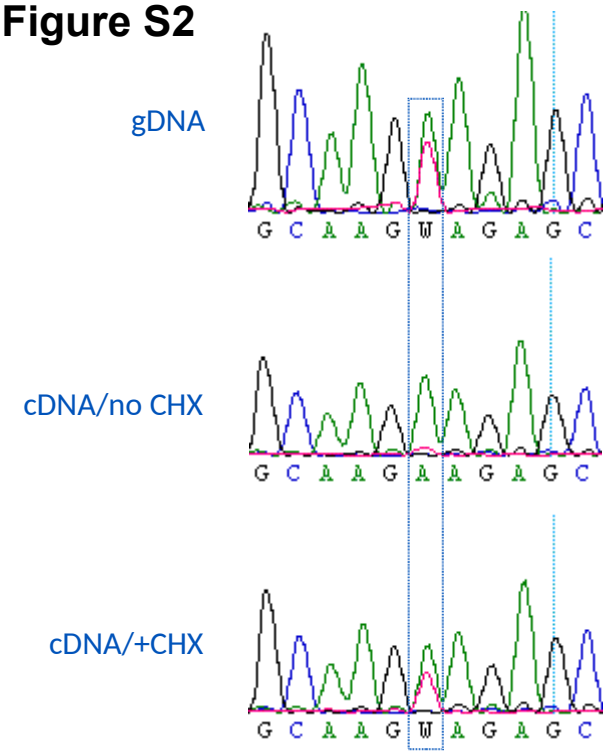


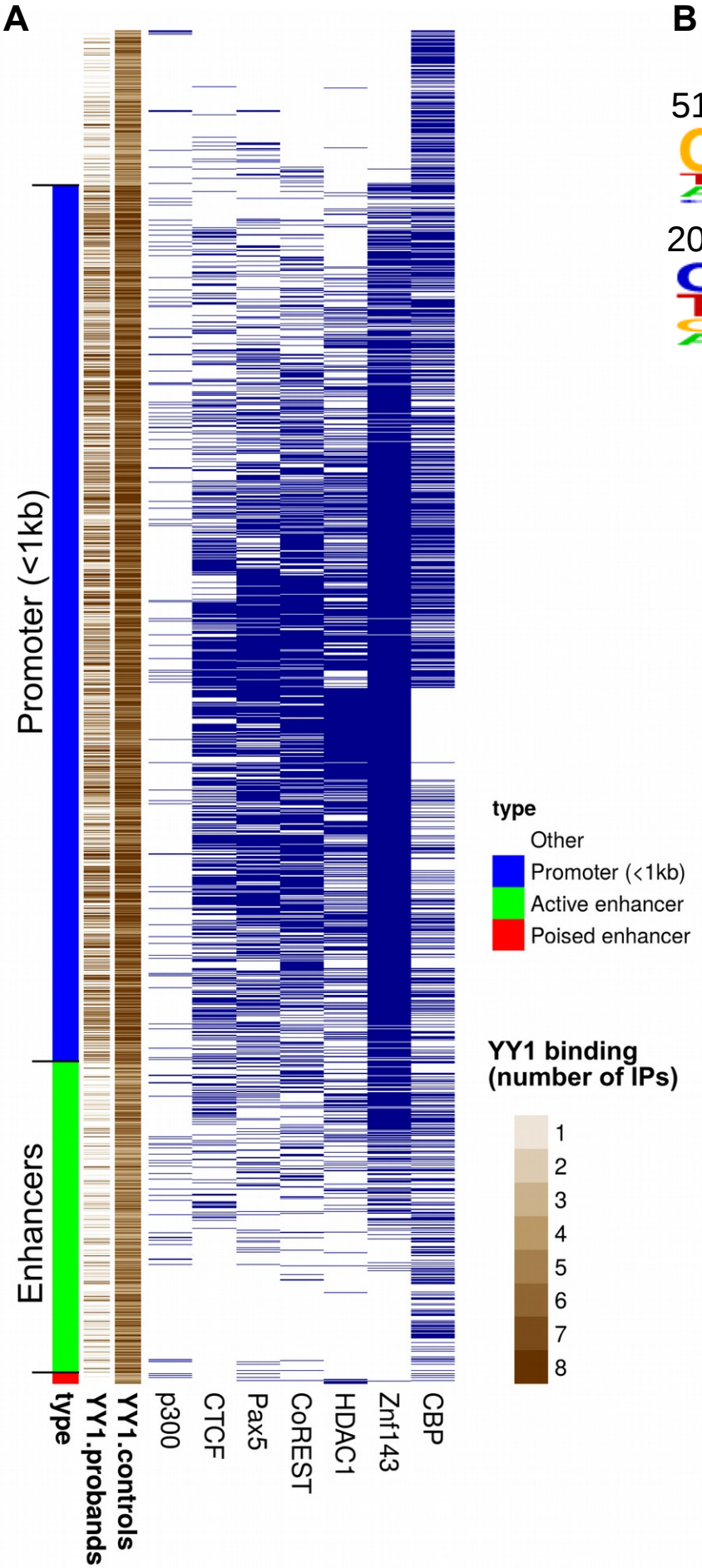
Figure S2: YY1 mRNA fom patient carrying mutation p.Lys179* is degraded in LCLs by Non-sense Mediated Decay (NMD). Sequence chromatograms showing normal and mutant alleles in genomic DNA (gDNA), the absence of the mutant (in absence of cycloheximide treatment) and the presence of the normal and mutant cDNA (in presence of 100ng CHX/6hr that inhibits NMD) in the affected LCLs. Normal/mutant sequence is boxed.

Figure S3

Sample	Antibody	# peaks	By sample		By condition	in all
Control1	1	4023	3198	2200	1932	1026
Control1	2	3807				
Control2	1	2320	2314			
Control2	2	4604				
Control3	1	3921	2678	2613		
Control3	2	2802				
Control4	1	7307	5728			
Control4	2	6940				
Control5	1	3178	2542	2542		
Control5	2	2908				
p.Asp380Tyr	1	3154	2189	1260	1029	
p.Asp380Tyr	2	2217				
p.Lys179*	1	1505	1285			
p.Lys179*	2	1347				
Deletion	1	1286	1094	1053		
Deletion	2	1146				
p.Leu366Pro	1	1760	1270			
p.Leu366Pro	2	1289				

Figure S3: Overlap of YY1 ChIPseq peaks across IPs.

Figure S4



B **DNA Motifs enriched in regions occupied by YY1**

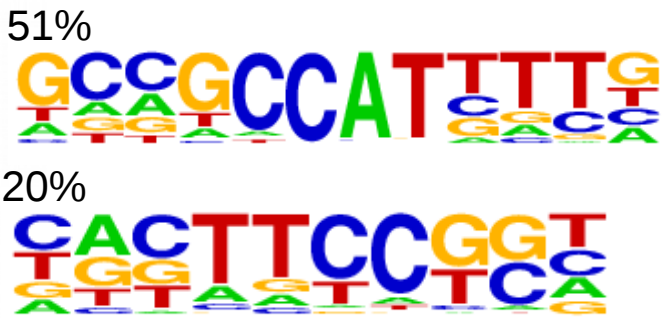


Figure S4: Enriched motifs and co-factor bindings in YY1 peaks. **A:** Overlap of YY1 peaks with relevant co-factors; each horizontal line represents a YY1 peak, shown in blue if it overlapped with a given factor. **B:** De novo motif found enriched in regions occupied by YY1, along with the percentage of target regions.

Figure S5

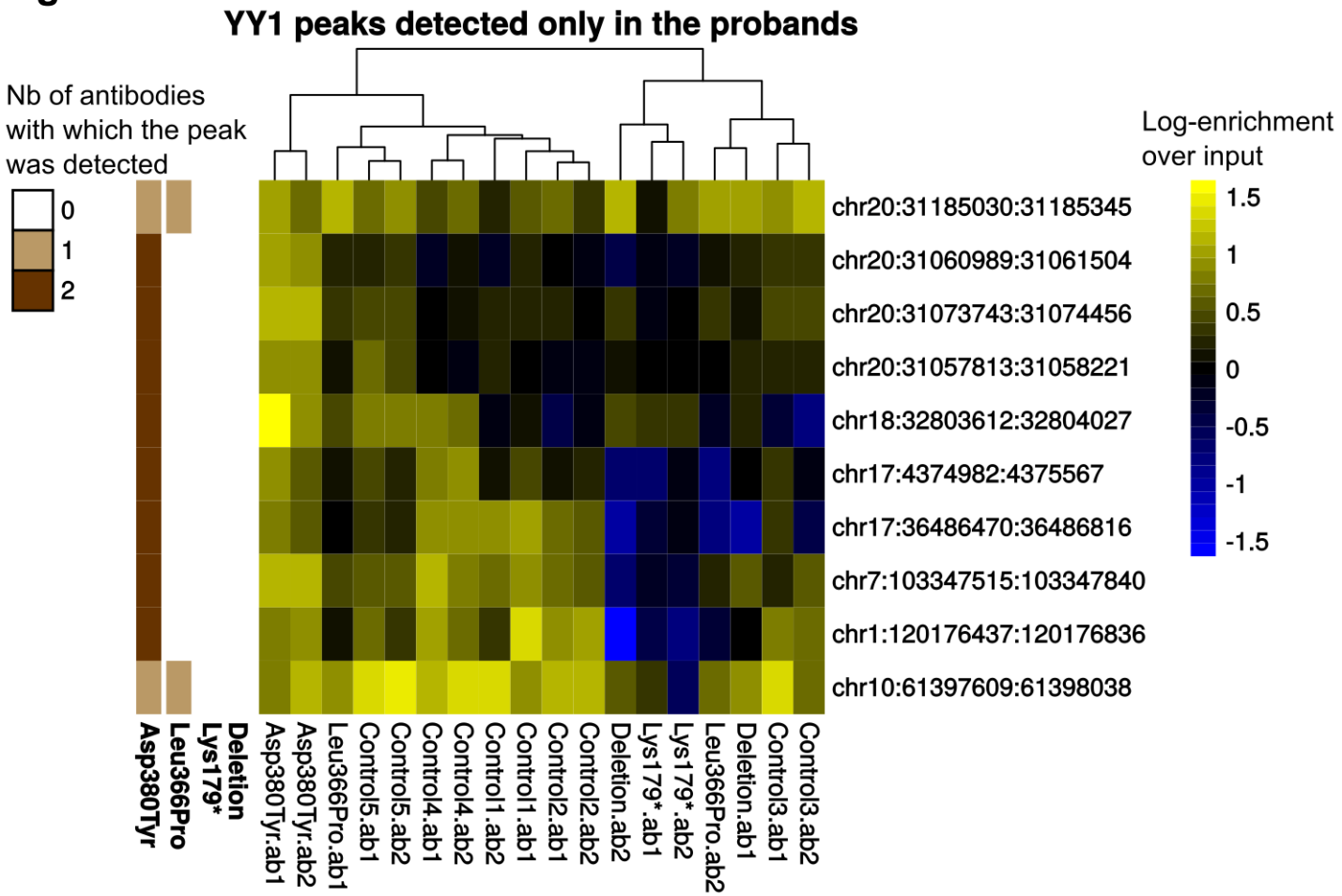


Figure S5: YY1 peaks detected only in the probands. YY1 peaks detected only in the probands nevertheless all show enrichment in nearly all control IPs.

Figure S6

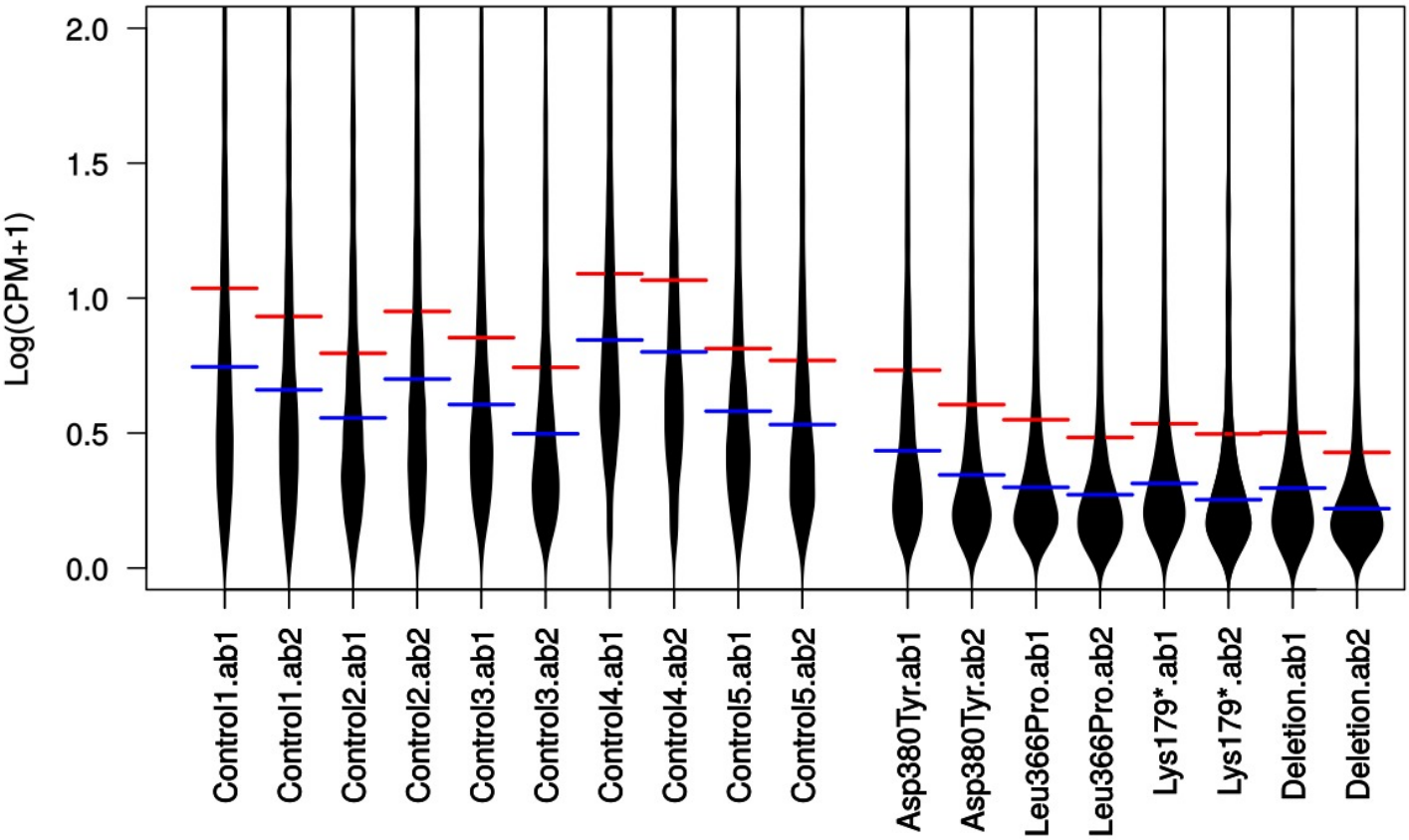


Figure S6: Distribution of relative YY1 counts (reads per million mapped reads) in the union of YY1 enriched regions in each IP. Blue lines represent medians, while red lines represent averages. The difference between means and medians are statistically significant under an Mann-Whitney test (p~4.6e-5).

Figure S7

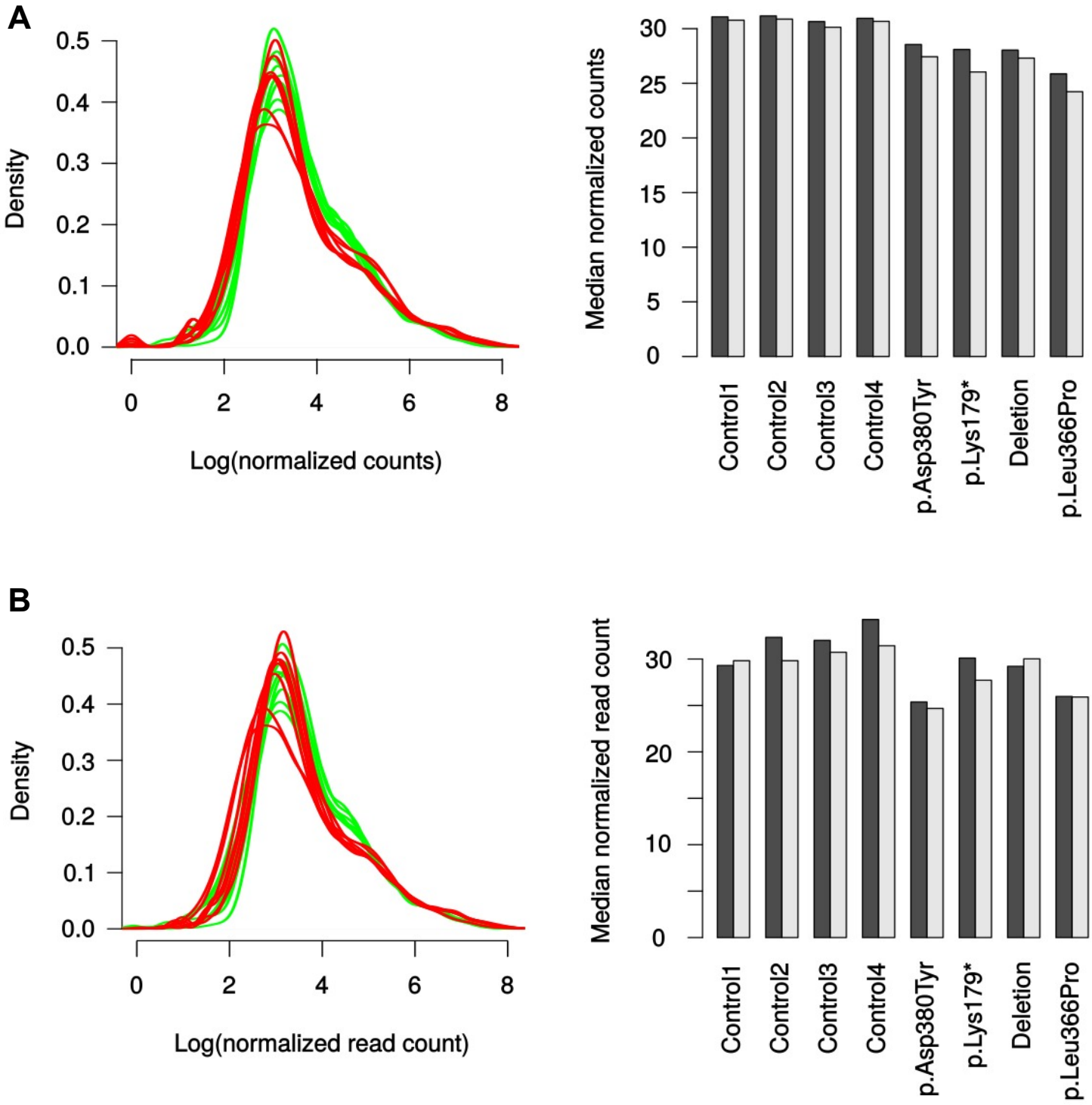


Figure S7: Global loss of YY1 binding confirmed by alternative normalization methods. A: Trimmed mean of M-values (TMM), p~ **B:** Normalization on conserved peaks (peaks detected across all samples). The left panels show the distribution of normalized read counts (with proband IPs in red), and the right panel shows the median of YY1-enriched regions (the two colors representing the two antibodies). The two shades of grey indicate the two antibodies.

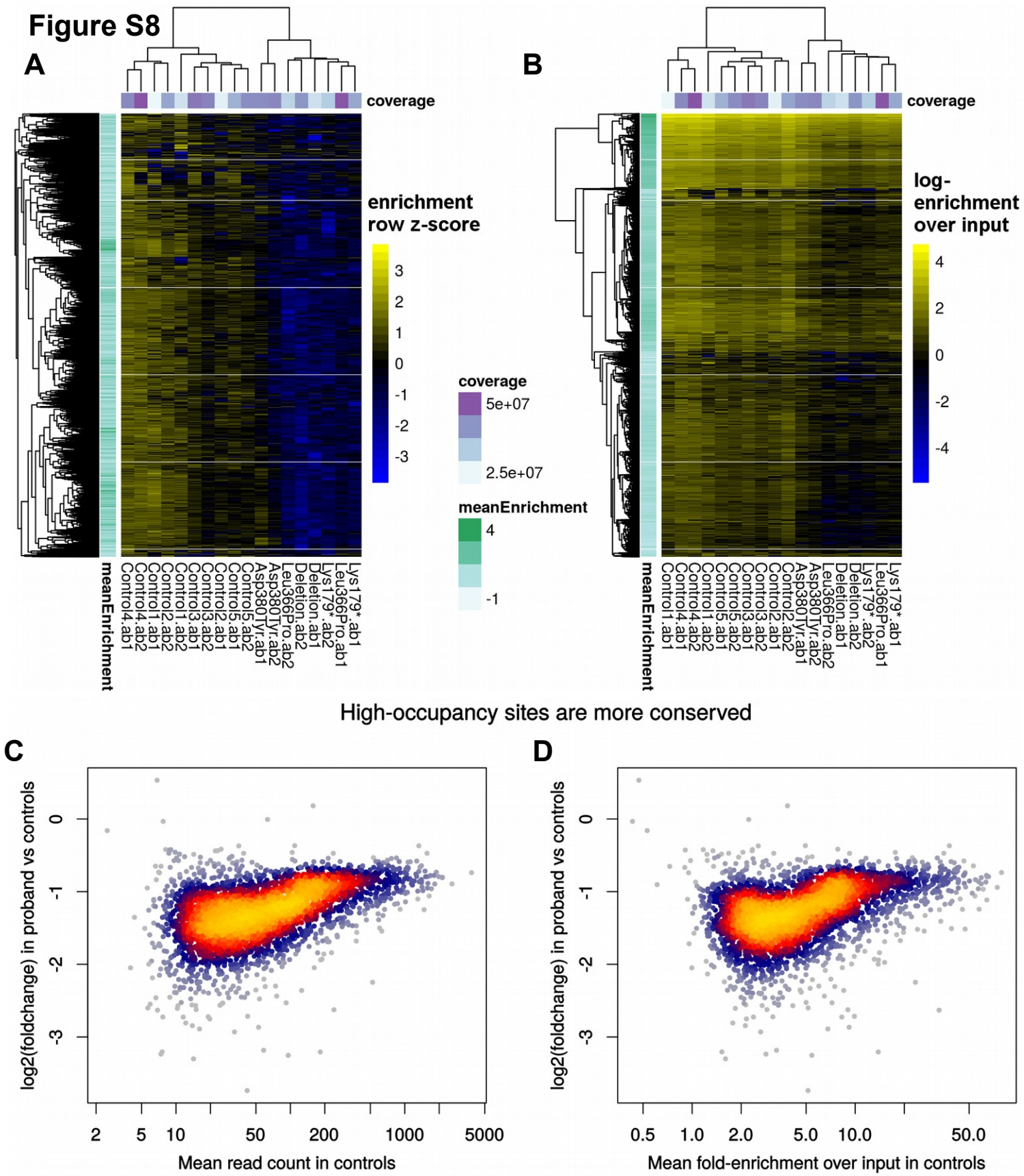


Figure S8: Quantitative analysis of YY1-bound regions. A: Heatmap of the enrichment z-scores for each peak across the union of YY1-enriched regions. **B:** Log-enrichment over input across all YY1-enriched regions; highly-enriched regions show a smaller decrease in the probands. **C:** The foldchange in YY1 signal is stronger in regions with lower read count. **D:** The foldchange in YY1 signal is stronger in regions with lower enrichment over the input. Read counts were normalized to library size for calculating enrichments and foldchanges.

Figure S9

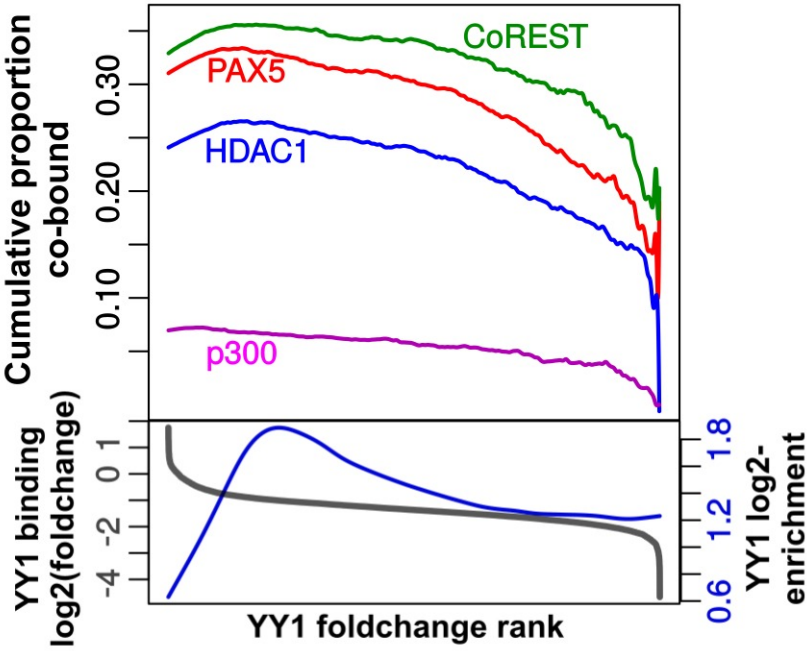


Figure S9: Cofactor co-occupancy across YY1 binding foldchanges. A: YY1 peaks were first ranked according to their foldchange in proband samples (represented with the grey line in the lower panel). Then, for each point in this ranking, the proportion of regions below this rank that are overlapping the respective cofactor was calculated and plotted. We can observe that YY1 peaks showing a more important decrease in the probands (right hand side) are comparatively less enriched for these co-factors than less affected YY1 bindings.

Figure S10

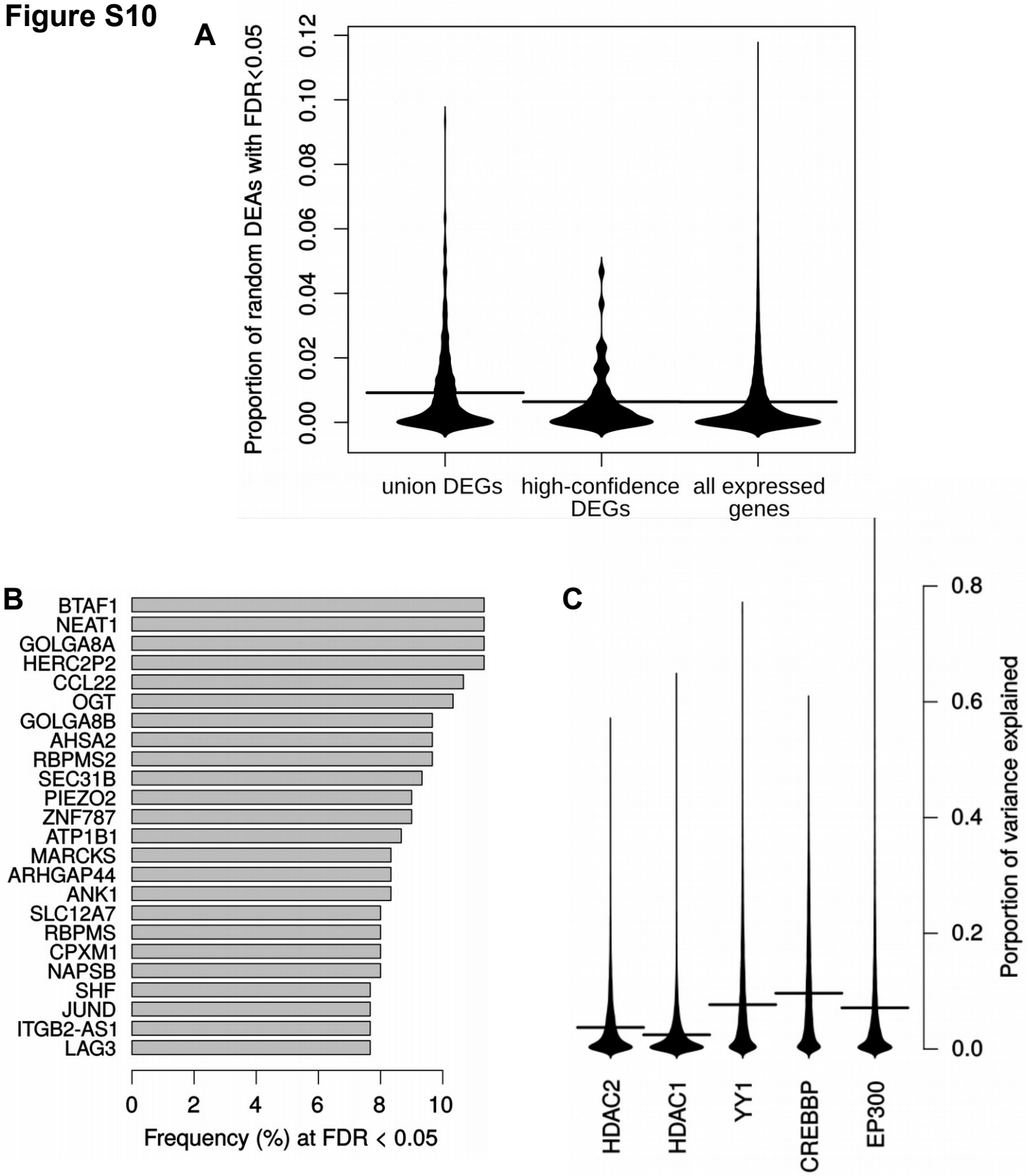


Figure S10: Results of 300 sex-balanced random differential expression analysis (3 vs 3 samples) across the HapMap LCL dataset. A: The distribution, across genes (either the DEGs of this study or all expressed genes), of the frequency at which they were found differentially-expressed in random comparisons. **B:** The most recurrently differentially-expressed genes across random analyses. **C:** Distribution of gene-wide transcriptional variance explaied by variations in YY1 and YY1-related factors.

Figure S11

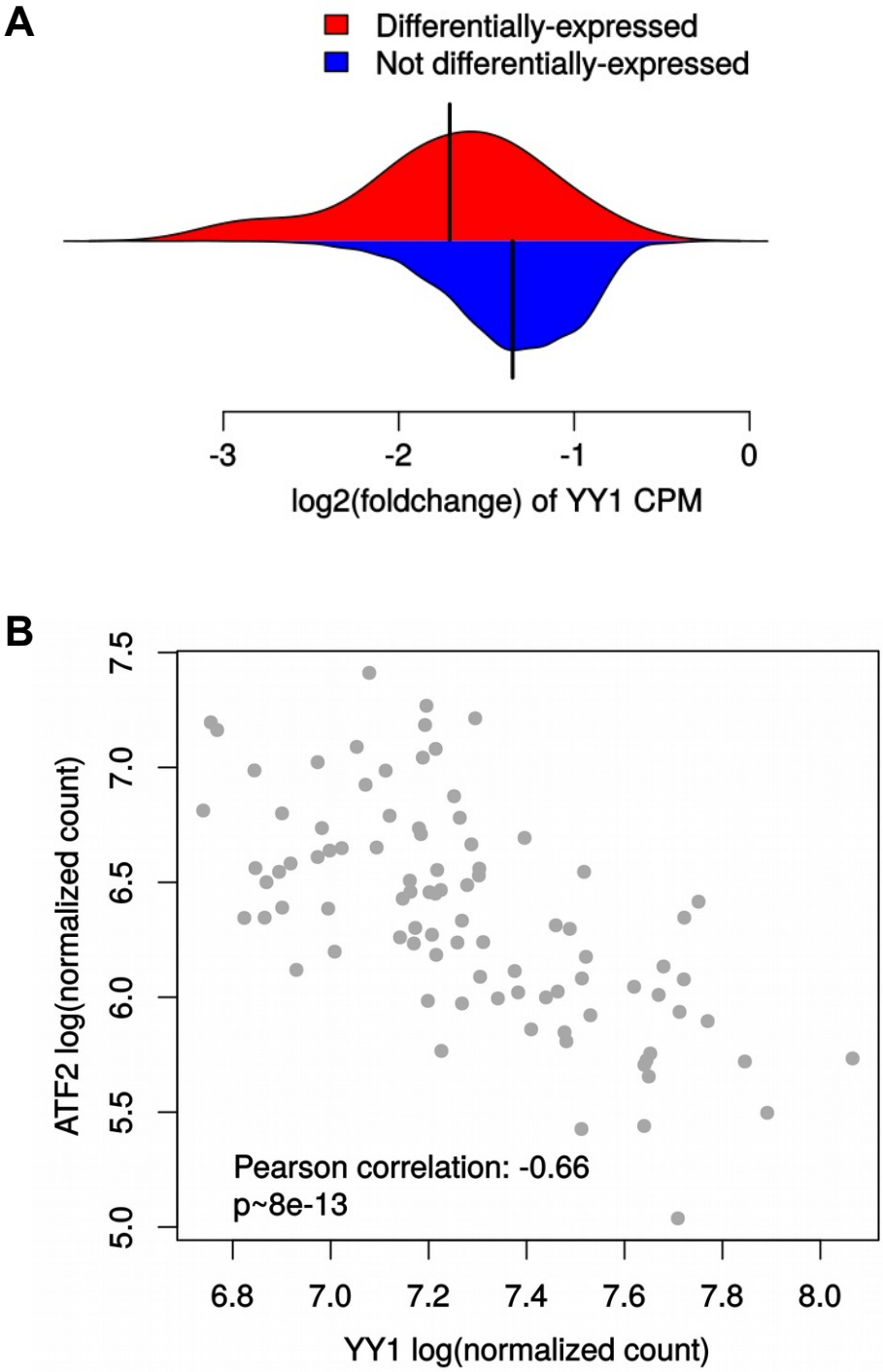


Figure S11: Correlation of YY1 foldchange with DEGs and ATF2 expression.
A: Distribution of YY1 binding foldchanges (in Counts per Million Reads mapped, CPM) at the TSS of differentially-expressed versus not significantly differentially-expressed genes. The difference is statistically significant (p~1e-4), indicating that the change in YY1 tends to result in transcriptional change, although the large overlap between the two distributions indicates that other factors distinguish the dosage-sensitive from dosage-insensitive targets.
B: ATF2 expression is significantly associated with YY1 mRNA levels across all LCL transcriptomes used in this study (18 samples of the internal dataset + HapMap LCLs).

Figure S12

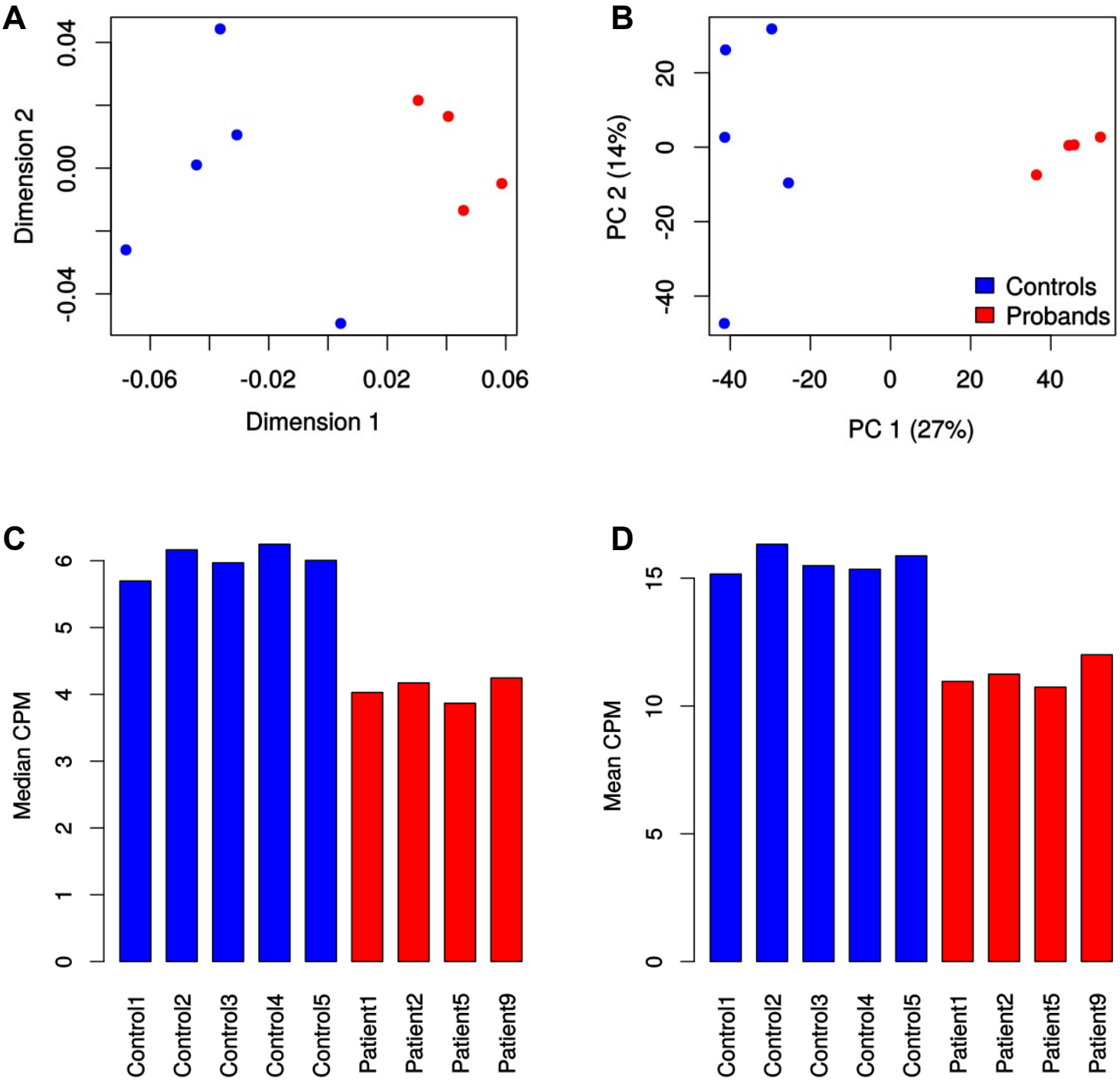


Figure S12: ChIPseq for H3K27ac in LCLs. A: Multi-dimensional scaling of the read counts across the union of enriched sites, based on distances calculated from the inverse of Pearson's correlation (and hence not relying on the assumptions of the linear normalization). **B:** Principal component analysis of the read counts (based on Euclidean distances) after linear normalization. **C-D:** Median (**C**) and mean (**D**) Counts Per Million reads mapped for each sample.

Figure S13

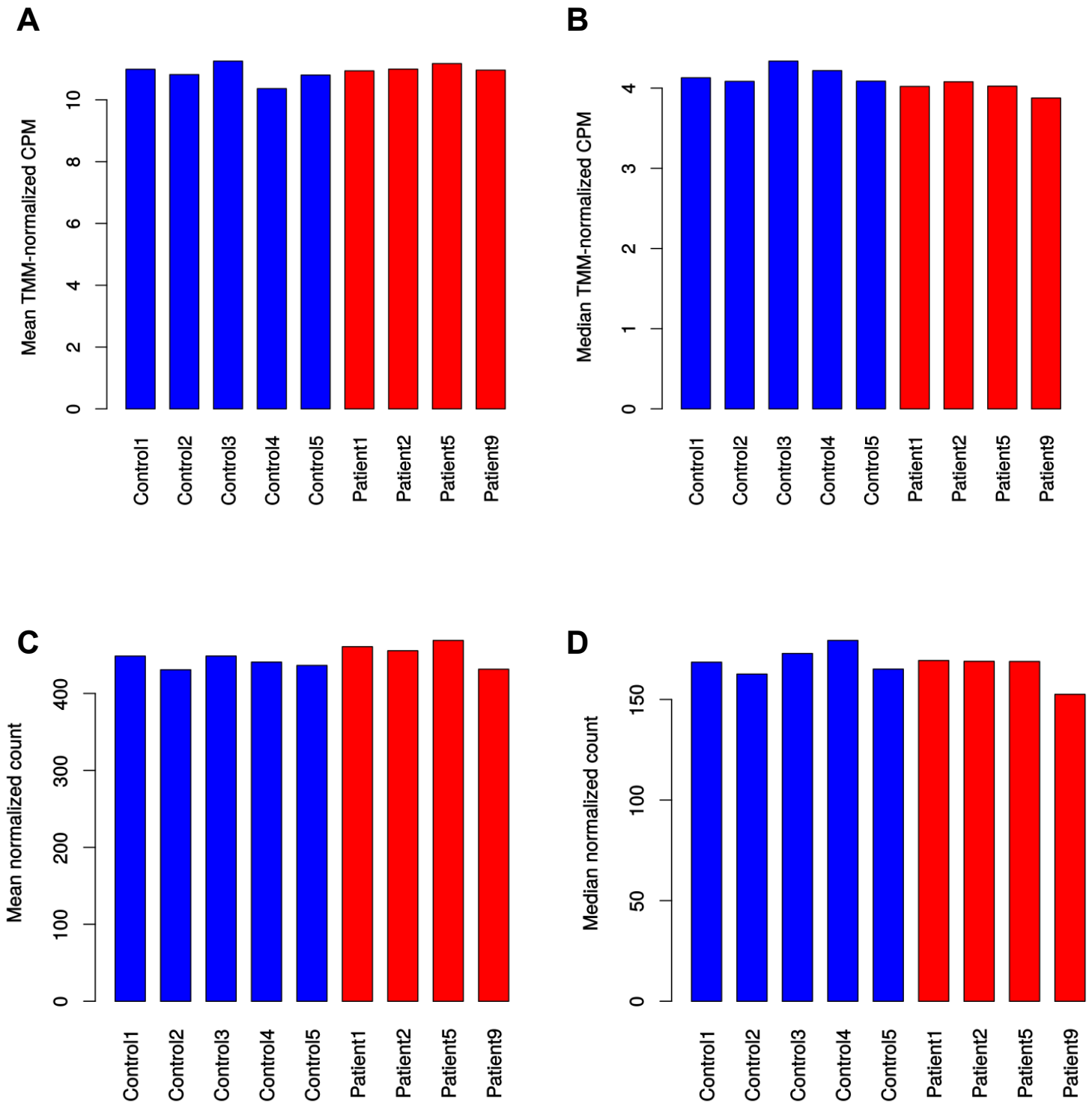


Figure S13: The global decrease in H3K27ac could not be corroborated by alternative normalization methods. A-B: Mean (**A**) and median (**B**) TMM-normalized counts per million reads sequenced. **C-D:** Mean (**C**) and median (**D**) counts in the union of H3K27ac regions linearly normalized on the peaks common to all samples.

Figure S14

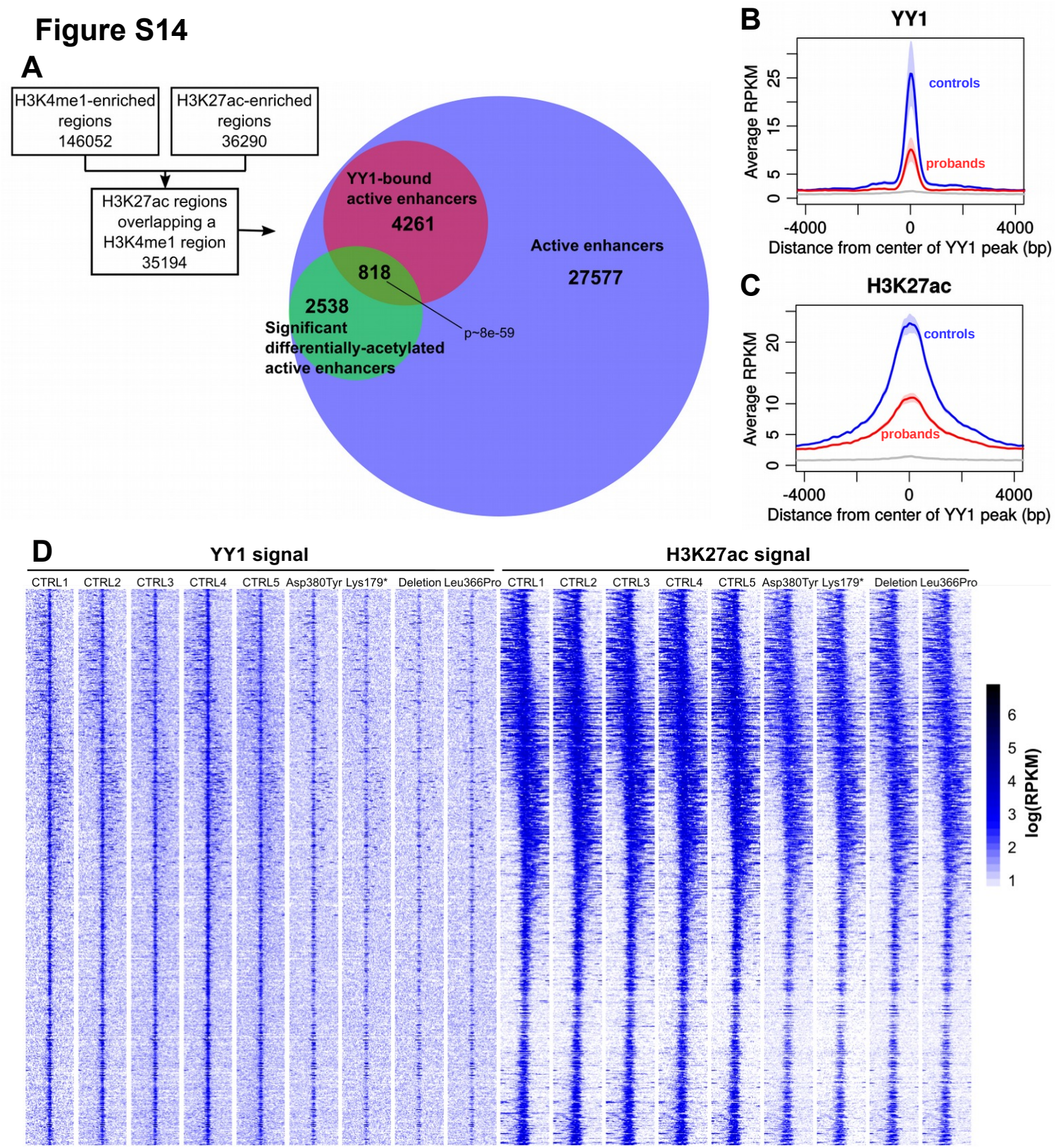


Figure S14: YY1 and H3K27ac at active enhancers. A: We first defined a set of active enhancers using the H3K27ac regions overlapping H3K4me1 regions, and considered which of those enhancers were differentially H3K27ac (under the assumption of no global change in H3K27ac) and which were overlapping a YY1 peak. Of note, when specifically quantifying H3K27ac on YY1 peaks and using library size normalization, 82% of YY1-bound enhancers show statistically significant decrease in H3K27ac. **B-C:** YY1 (**A**) and H3K27ac (**B**) average density at YY1-bound, differentially-acetylated enhancers. **D:** Read density (in log-RPKM) at each YY1-bound, differentially-acetylated enhancer, +/- 5kb around the center of the YY1 peak. RPKM stands for Reads Per Kilobasepair per Million reads mapped.

Figure S15

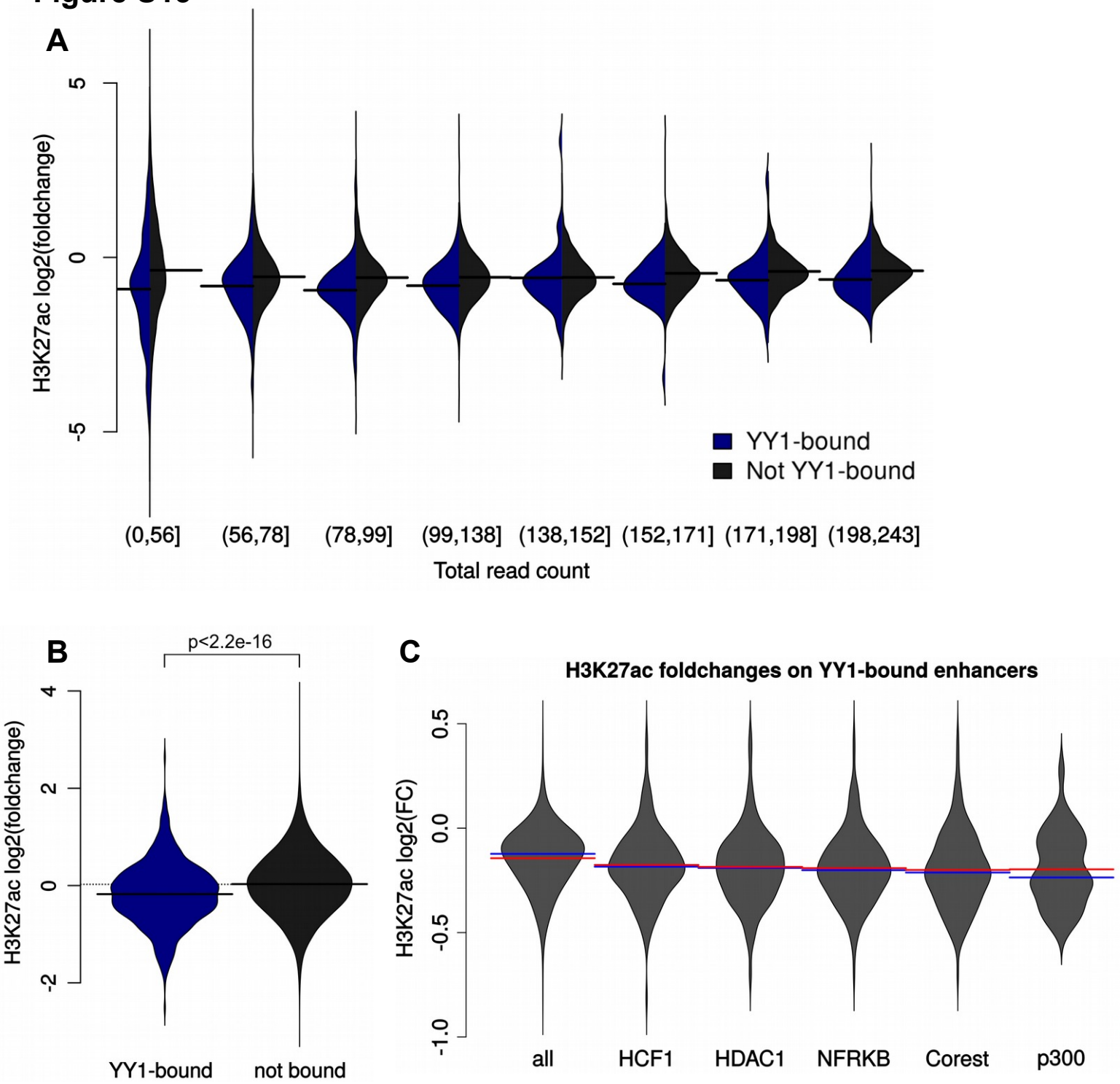


Figure S15: Preferential loss of H3K27ac in YY1-bound enhancers. A: The preferential loss of H3K27ac at YY1-bound versus not-YY1-bound active enhancers is independent of the total H3K27ac read count, and hence of potential differences in signal to noise ratio. **B:** The preferential loss at YY1-bound enhancer can be reproduced using TMM normalization (which assumes no global difference) instead of library size normalization (as in Figure 5C). The p-value indicated is the result of a two-tailed t-test. **C:** Distribution of H3K27ac foldchanges across subset of YY1-bound enhancers bound by cofactors. Blue lines represent median, red lines represent means. H3K27ac loss at YY1-bound enhancers does not seem to be attributable to any particular chromatin remodeling complex with which YY1 interacts.

Figure S16

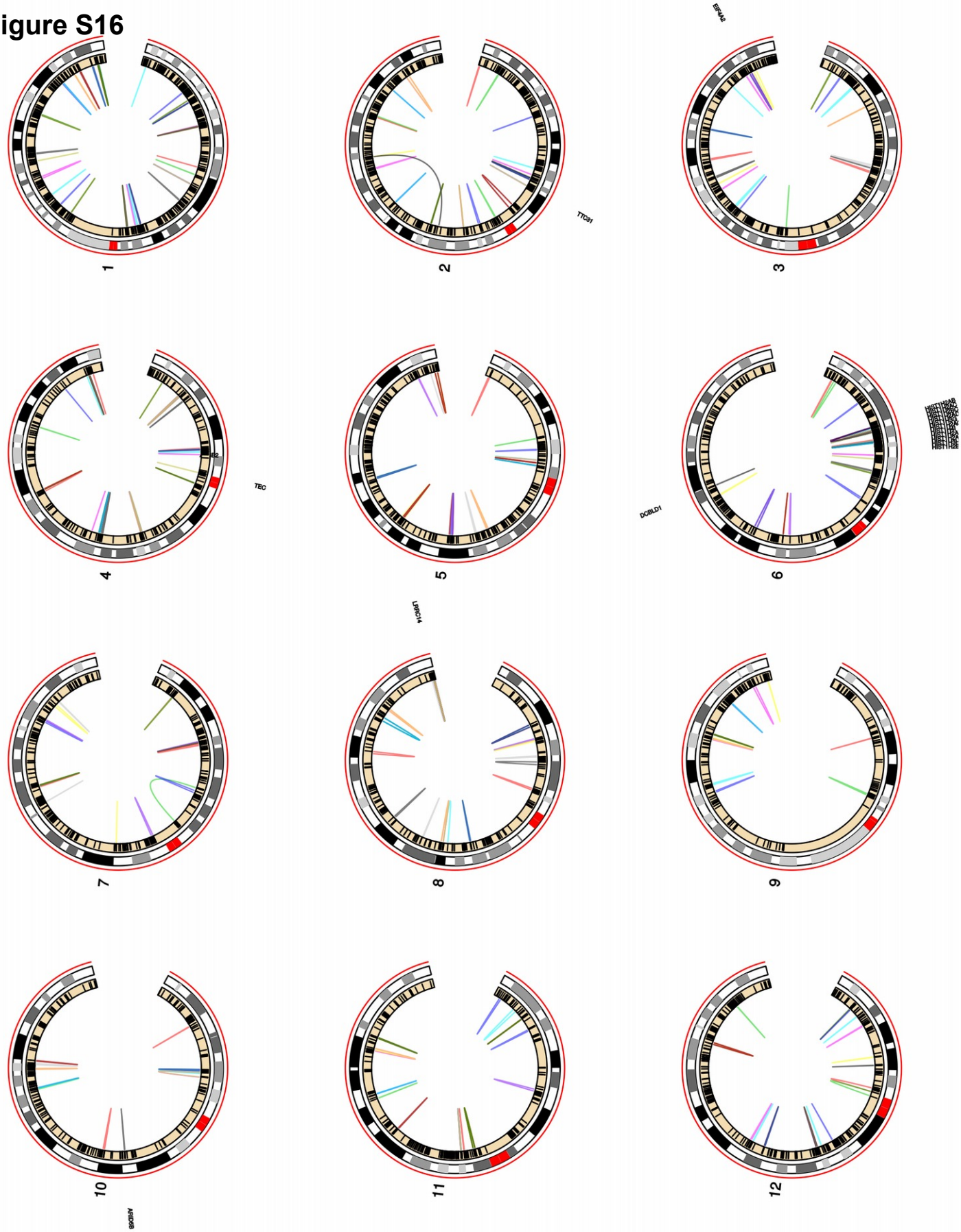


Figure S16: Distal YY1 bindings for chromosomes 1 to 12. For each chromosome, the 3D chromatin interactions (derived from GM12878 LCLs) connecting YY1 bindings to target genes are shown as arcs in the middle, and the YY1 binding sites are shown as bars below the ideograms. Differentially-expressed genes with a distal YY1 binding are indicated on the outside.

Figure S17

Figure S17: Distal YY1 bindings for chromosomes 13 to Y. For each chromosome, the 3D chromatin interactions (derived from GM12878 LCLs) connecting YY1 bindings to target genes are shown as arcs in the middle, and the YY1 binding sites are shown as bars below the ideograms. Differentially-expressed genes with a distal YY1 binding are indicated on the outside.

Figure S18

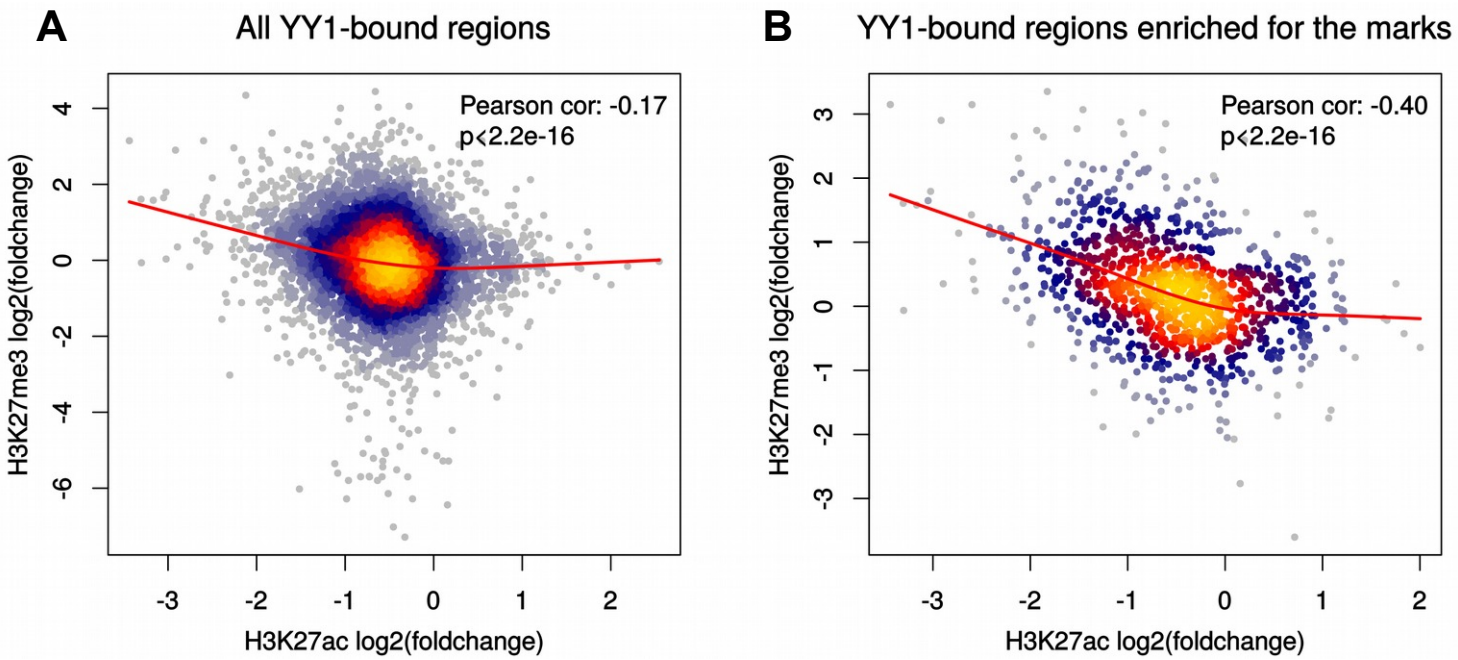


Figure S18: Across YY1-bound regions, the loss of H3K27ac is associated with an increase in H3K27me3. A: Density plot of foldchanges in H3K27ac and H3K27me3 across all YY1-enriched regions. **B:** Density plot of foldchanges in H3K27ac and H3K27me3 across YY1-bound regions that show an enrichment for H3K27me3 and H3K27ac respectively in at least one sample. The red line shows a smooth spline fitted on the data.

Table S1

Identifier	Chr start (hg19)	Chr end (hg19)	Size (Mb)	Inheritance	UPD(14) cluster	Allele	Other chr aberrations	Phenotype
RUMC1	88550163	101828366	13.3	de novo	yes	NA	-	IUGR, hypotonia, F, hearing loss, D, recurrent infections
266323	96723247	100938073	4.2	unknown	no	-	dup chr18:27,024,505-28,465,203	ID, IUGR, M, hypotonia, F, malrotation, micropenis, hypoplastic scrotum, C, D
249213	98099681	107268432	9.2	de novo	yes	NA	-	IUGR, patent ductus arteriosus, cryptorchidism, tethered cord, small chest, D
254518	98534687	101064183	2.5	de novo	no	-	-	ID, SGA, SS, hypotonia, spasticity, micropenis, C, F, strabism, nystagmus, D, recurrent infections
SA1	99200000	100900000	1.8	de novo mosaic	no	-	-	moderate ID, atrial septal defect, C, cleft palate, ureteral duplication, VUR, F, GER, hypermetropia, D
250486	99357235	103249889	3.9	de novo	yes	NA	-	severe ID, IUGR, hypotonia, microcephaly, hypospadia, cryptorchidism, mild ventriculomegaly, F, D
SA2	99700000	101200000	1.5	*de novo*	yes	NA	-	moderate ID, hypotonia, corpus callosum hypoplasia, C, D
263205	100260674	105692495	5.4	*de novo*	yes	Mat	-	fetal hygroma colli, partial esophageal atresia, club feet, tricuspid insufficiency, right ventricle thickness, TOP
263711	100397065	101502752	1.1	de novo	yes	Pat	del chr8:3,623,048-3,659,817	ID, IUGR, hypotonia, atrial and ventricular septal defect, hypoplastic aortic arch, duplex kidney, F, D, recurrent infections
256842	100508222	100771805	0.3	*de novo*	no	-	del chr9:139,242,169-139,472,515 *de novo*	mild ID, SGA, SS, hypotonia, hyperactivity, F, hypermetropia, D
267668	100508222	102235753	1.7	*de novo*	yes	NA	-	corpus callosum agenesis, enlarged kidneys, SUA, TOP
272547	100716801	100791356	0.1	*de novo*	no	-	dup chr20:7,287,394-7,761,372 inherited	mild DD, very active, poor concentration, strabism, F, long slim hands and feet, deviated toes, constipation, D
271459	100735192	101126789	0.4	*de novo*	no	-	-	moderate ID, SGA, strabism, enlarged lateral ventricle, white matter changes, F, long halluces, D

Table S1: Patients with deletions encompassing *YY1*

C: craniosynostosis, D: dysmorphisms, DD: developmental delay; F: feeding problems; GER: gastro-oesophageal reflux, ID: intellectual disability; IUGR: intrauterine growth retardation; M: microcephaly; SGA: small for gestational age; SS: short stature; SUA: single umbilical artery; TOP: termination of pregnancy; VUR: vesicoureteric reflux

Table S5

Sample		Western Blot	RNAseq		ChIPseq			
Sample	Mutation	YY1	1st	2nd	YY1 (sc-1703)	YY1 (sc-281)	H3K27ac	H3K27me3
Control1		X	X	X	X	X	X	X
Control2		X	X	X	X	X	X	X
Control3		X	X	X	X	X	X	X
Control4		X	X	X	X	X	X	X
Control5		X		X	X	X	X	X
Control6				X				
Patient1	p.Asp380	X	X	X	X	X	X	X
Patient2	p.Leu366	X	X	X	X	X	X	X
Patient5	p.Lys179*	X	X	X	X	X	X	X
Patient11	Deletion	X	X	X	X	X	X	X

Table S5: Summary of the experiments

Author Contributions

ATVvS, DAK, GT and BBAdV conceived the project and designed the study. ATVvS, KK, TT, AR, KS, EF, DD, AH, PS, ZP, JAR, DRB, LF, CPS, JF, PT, SDM, KMN, YS, WJC, HGB, JHMSH, BWMvB, WMN, LELMV, DAK, and BBAdV were involved in recruitment of individuals with *YY1* mutations. ATVvS, JM, EH, JA, ASP, PC, SP, SAL, SK, PMT, CBA, AR, AvH, RP and DAK were involved in recruitment of individuals with deletions of *YY1.* SL and CG performed the enrichment calculation. JG and RK evaluated the LCLs of individuals for mRNA and protein levels and performed test for NMD. MG performed the transcriptomic and epigenomic experiments on individuals' samples. PLG and AV performed the computational analysis of transcriptomic and epigenomic data. AV performed the modeling of YY1 mutations on the crystal structure. MG, ATVvS, PLG, DAK, GT and BBAdV prepared the first draft of the manuscript. All authors contributed to the final manuscript.