TABLE OF CONTENTS

SUPPLEMENTARY DATA

Supplementary data are available as separate files.

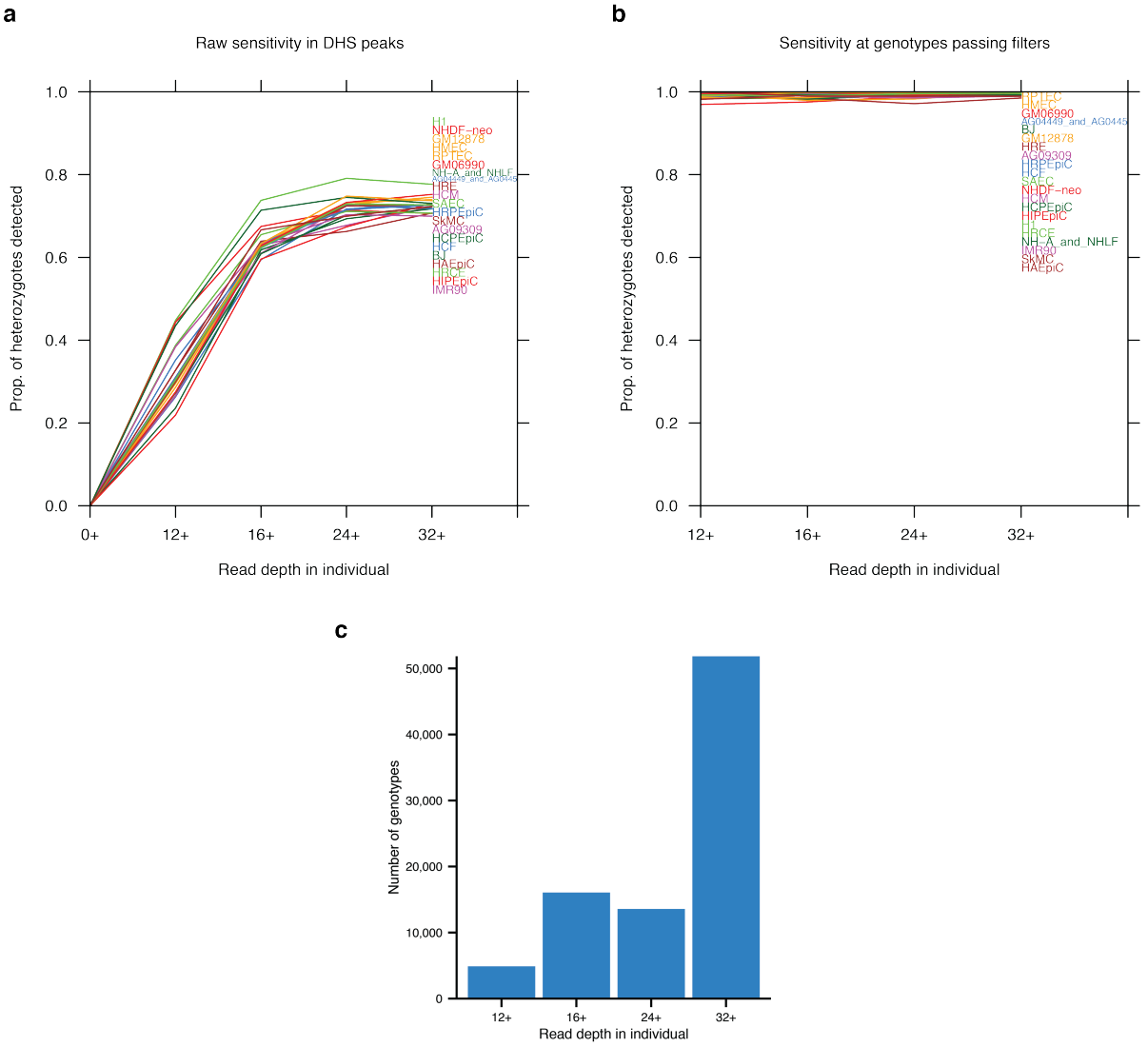
Supplementary Data 1. SNPs tested for imbalance in DNA accessibility.
SNPs are listed by their hg19 coordinates. The rsid is used for SNPs in dbSNP138. SNPs are classified as imbalanced as in **Fig. 1c**. PctRef refers to the proportion of reads mapping to the reference allele (**Fig. 1d**).

Supplementary Data 2. TF clusters of similar motifs.
Motif weblogos from JASPAR, UniPROBE, and Jolma et al. 2013[35] databases grouped by TF cluster. Motifs from TRANSFAC are listed by name without showing weblogo.

Supplementary Data 3. SNVs predicted to affect DNA accessibility.
List of SNVs from dbSNP138 overlapping a TF recognition sequence in a DHS hotspot predicted to affect accessibility with a score greater than 0.10. The file is in extended bed format using hg19 coordinates and includes a header line. Each row contains the SNP coordinates and dbSNP ID, a score scaled as the probability of imbalance, the PWM name and strand orientation, the position of the SNP relative to the PWM match, and the two alleles of the SNP (relative to the strand of the PWM).

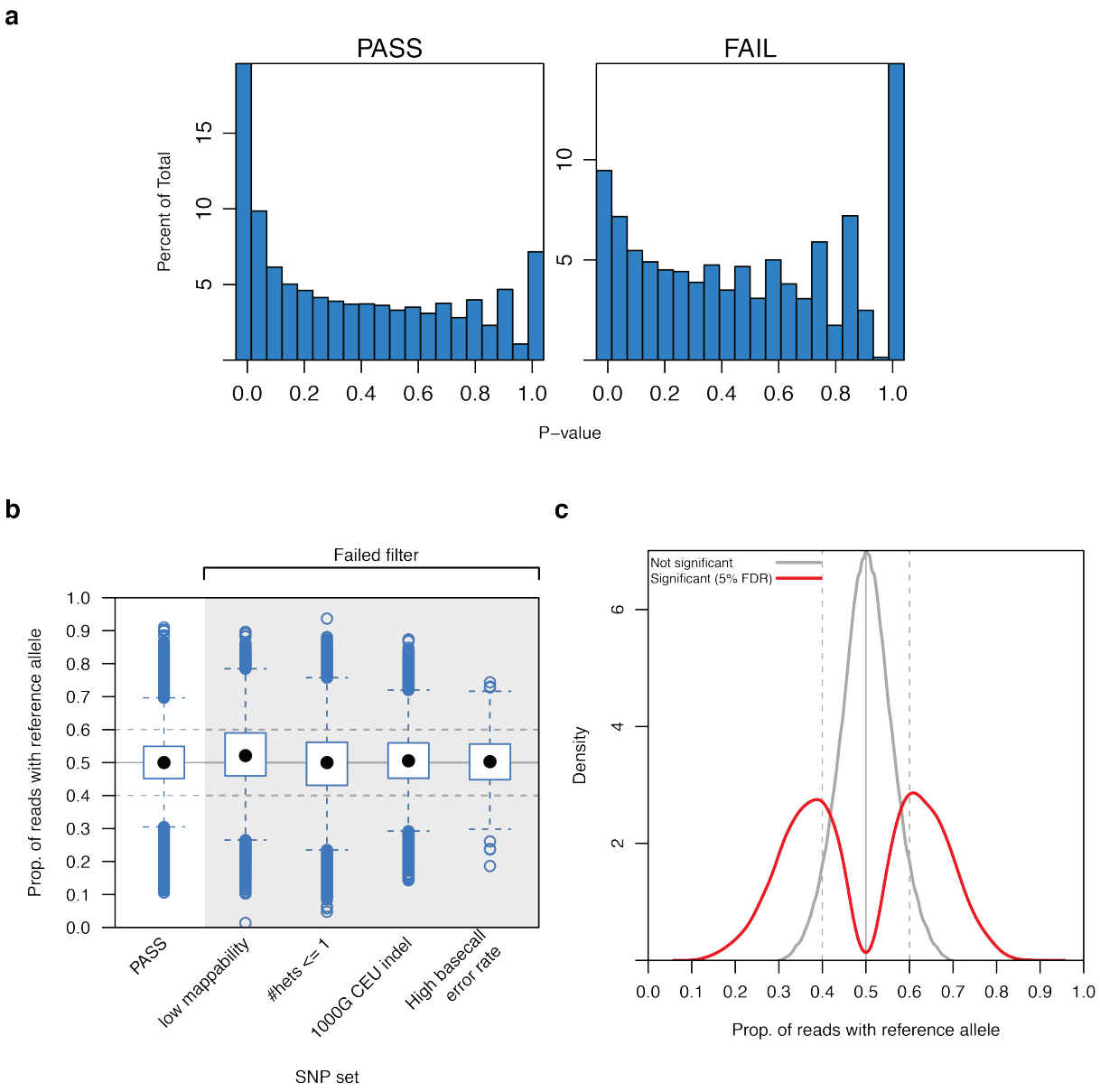
Nature Genetics: doi:10.1038/ng.3432

SUPPLEMENTARY FIGURES



Supplementary Fig. 1. Sensitivity of detection of heterozygous genotypes.
(a-b) Shown is the frequency of detection of known heterozygous sites from publicly available Illumina Human 1M-Duo genotyping array data for 23 individuals[22]. Values are averaged for individuals with replicate genotyping arrays. Note that we only called genotypes at sites with at least 12 reads. Shown are (a) raw sensitivity relative to all Illumina genotypes in DHSs; and (b) sensitivity for variant calls passing all filters (**Online Methods**). At high read depths (>32 reads) at variants passing all filters, we detect 99.4% of heterozygotes. (c) Counts of genotype calls for each read depth bin in **a-b**.

Nature Genetics: doi:10.1038/ng.3432

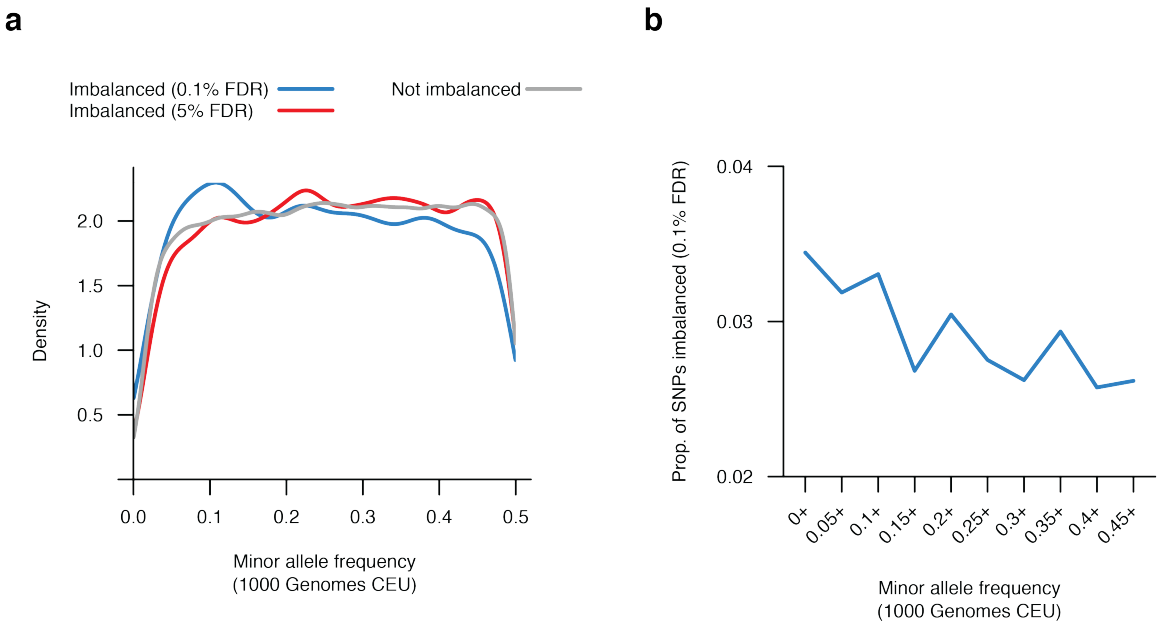


Supplementary Fig. 2. Identification of allelic imbalance.
(a) Binomial test *P*-values for variants passing all filters, or failing one or more filtering criteria (**Online Methods**).

(b) Allelic ratios relative to reference allele. Note the symmetry of 'PASS' variants around 0.5 and the reduced variance compared to the sets of the failed SNPs.

(c) Allelic ratios of SNPs passing filters, broken down by the presence of imbalance.

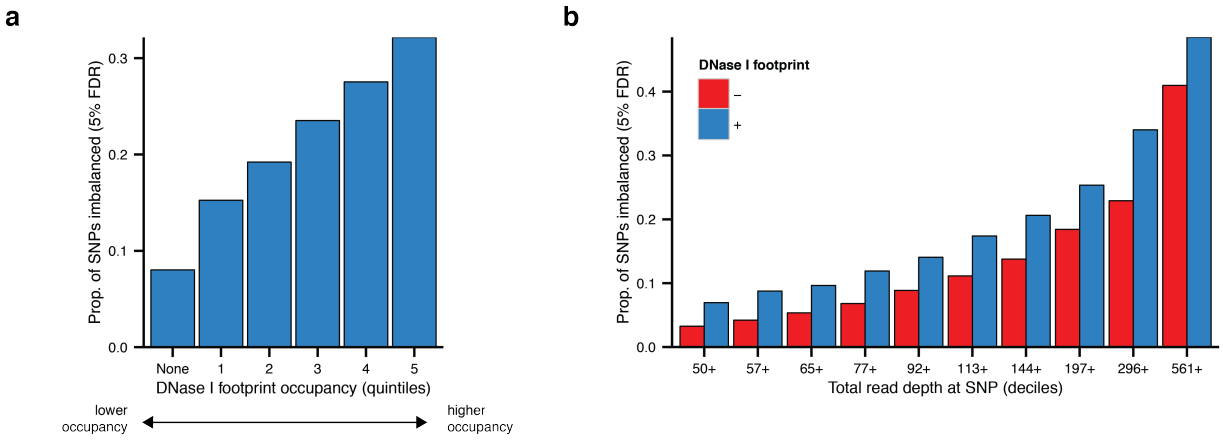
Nature Genetics: doi:10.1038/ng.3432



Supplementary Fig. 3. Imbalance and population allele frequency.
(a) Distributions of minor allele frequencies (MAF) from the 1000 Genomes CEU data for variants with and without imbalance in DNA accessibility. Only intergenic or intronic SNPs >1000 bp from a TSS were considered.

(b) Frequency of imbalance versus minor allele frequency.

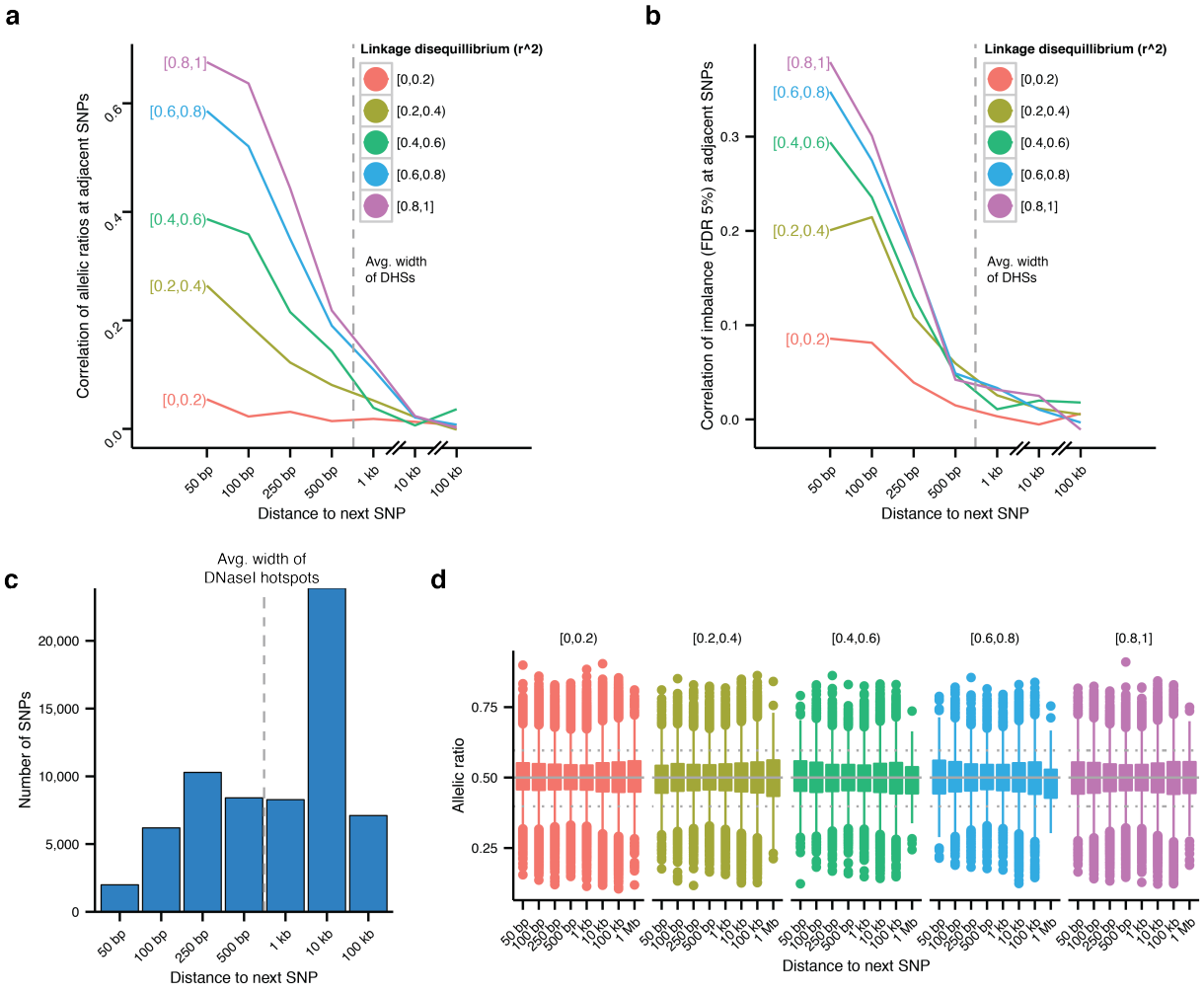
Nature Genetics: doi:10.1038/ng.3432



Supplementary Fig. 4. Imbalance quantitatively tracks DNase I footprint occupancy.
(a) Increased frequency of imbalance with higher occupancy indicated by DNase I footprint score.

(b) Increased frequency of imbalance in DNase I footprints occurs across all read depths.

Nature Genetics: doi:10.1038/ng.3432

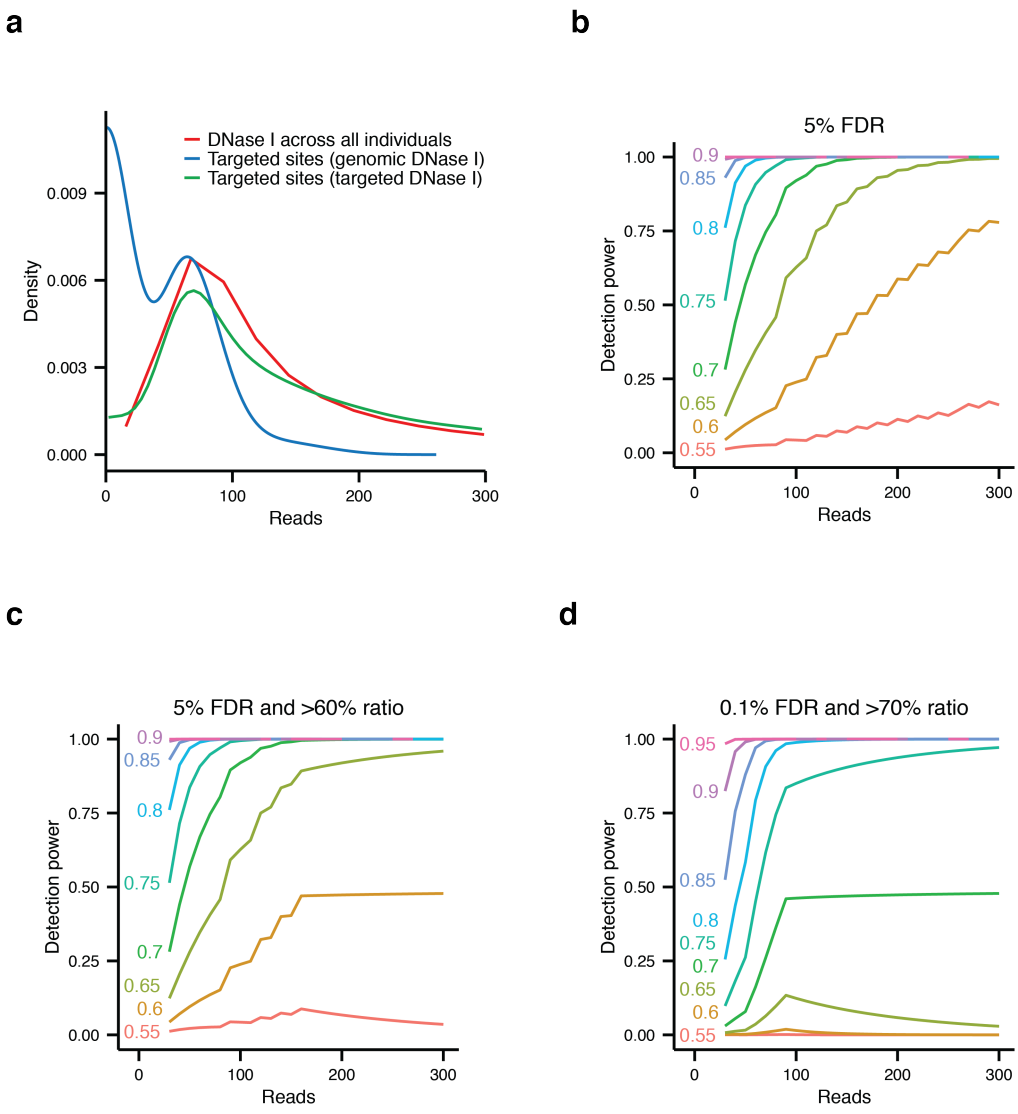


Supplementary Fig. 5. Linkage disequilibrium and correlation of imbalance of nearby sites. (a-b) Pearson correlation of allelic ratios (a) and presence or absence of imbalance (b) at adjacent SNPs broken down by distance to next SNP. Dashed line represents the median width of DHS hotspots overlapping SNPs in this study. SNPs are additionally broken down by linkage disequilibrium (r^2) between the pairs of SNPs in our samples.

(c) Number of SNPs in each bin from **Fig. 1e** (corresponding to the $r^2 \sim [0.8,1]$ bin in (a)).

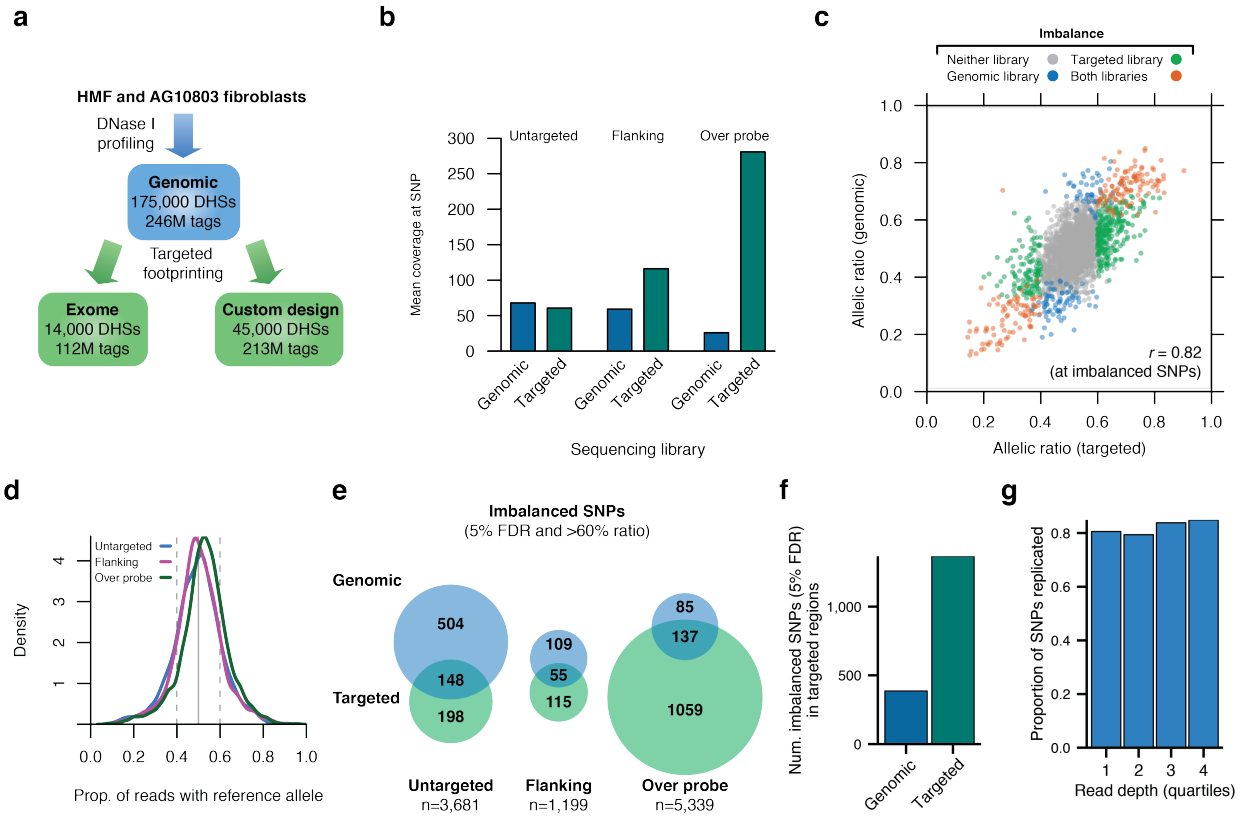
(d) Allelic ratios for SNPs broken down by proximity and linkage disequilibrium (r^2).

Nature Genetics: doi:10.1038/ng.3432



Supplementary Fig. 6. Calculated power to detect imbalance using the binomial distribution. (a) DNase I read depth for sites across all individuals and in targeted footprinting analysis. (b-d) Detection power versus read depth under the binomial is plotted for SNPs exhibiting various levels of imbalance. Lines are labeled with allelic ratios. Shown are the (b) FDR 5% threshold used for DNase I and ChIP-seq analysis; (c) 5% FDR and >60% imbalance threshold used for the targeted footprinting and cross-cell type analyses; and (d) 0.1% FDR and >70% imbalance threshold used for the strict set of imbalanced DNase I SNPs (**Online Methods**).

Nature Genetics: doi:10.1038/ng.3432



Supplementary Fig. 7. Targeted survey of imbalanced DNA accessibility.
(a) Summary of strategy and data used. Two genomic DNase-seq libraries were captured with two targeted footprinting capture designs each to augment coverage at a selected set of DHSs.

(b) SNPs were classified by location as: (i) not targeted, (ii) lying within flanking 250 bp of a capture probe, or (iii) directly overlapping the probe itself. Sequencing coverage increased specifically at targeted sites.

(c) Comparison of allelic ratios at SNPs overlapping probes in targeted footprinting and genomic libraries. Sites are colored by the presence of imbalance in one or both libraries. *r*, Pearson correlation. Only 1 of 192 SNPs that demonstrate imbalance in both targeted and genomic samples showed imbalance in opposite directions.

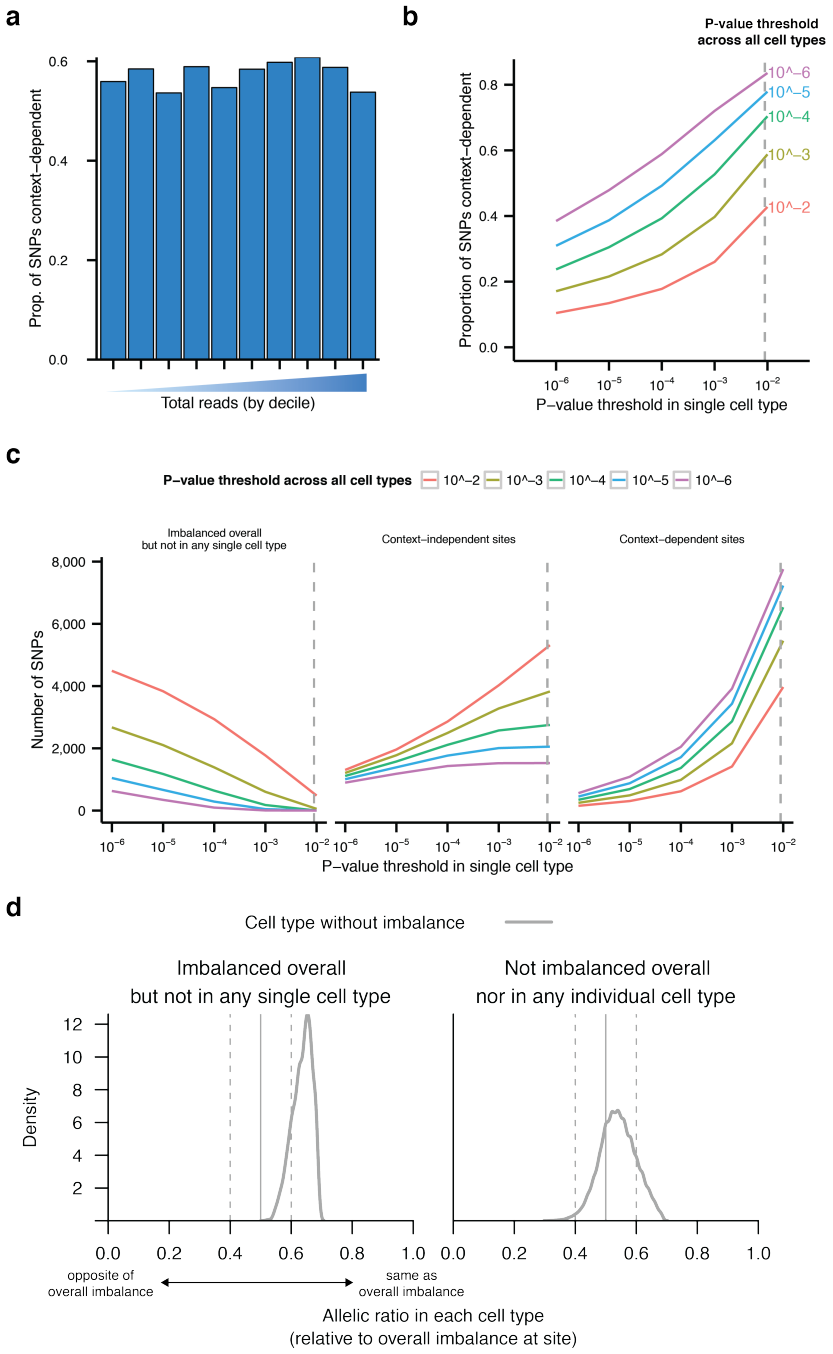
(d) Allelic ratios in capture data sets. Note the increase in bias for the reference allele specific to SNPs overlapping capture probes (green).

(e) Detection of novel sites of allelic imbalance in targeted libraries. Note that some strong sites afforded sufficient power to identify imbalance in genomic data or off-target regions.

(f) Counts of imbalanced variants identified in targeted and genomic libraries.

(g) Replication rate in the targeted data of SNPs from the full genomic data set of 114 cell types (measured as concordance of imbalanced/not imbalanced calls).

Nature Genetics: doi:10.1038/ng.3432

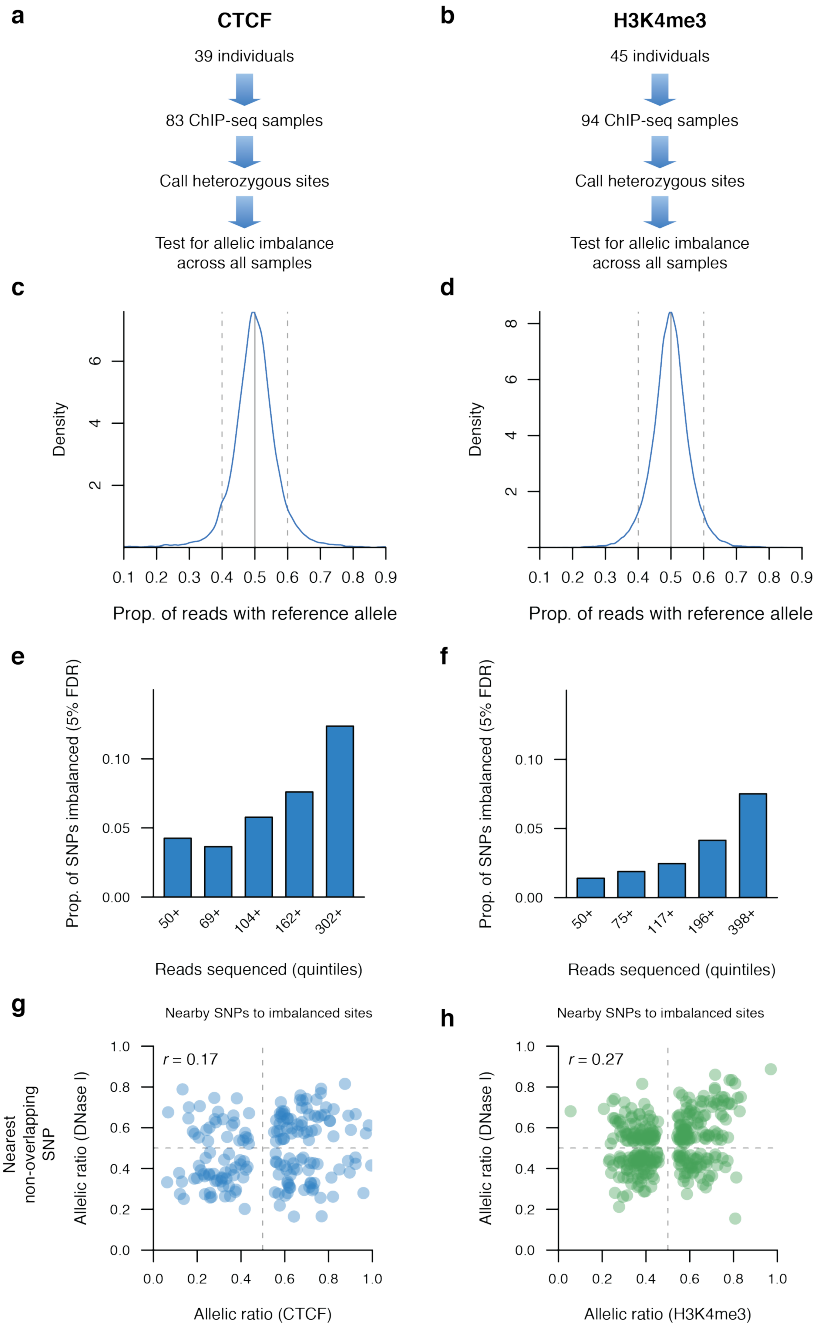


Supplementary Fig. 8. Cross-cell type analysis of imbalance.
(a-b) Proportion of SNPs with imbalance in a single cell type that are context-dependent (vs. context-independent) stratified by (b) total read depth across all cell types; and (c) P-value threshold for imbalance in single cell types and across all cell types.

(a) Counts of SNPs discovered at various P-value thresholds for imbalance in single cell types and across all cell types. A threshold of $P < 0.0089$ (matching the 5% FDR threshold from the global analysis) was used as the cutoff in a single cell type (dashed gray line) and across all cell types.

(d) Allelic ratios in samples without overall imbalance, oriented such that 1.0 represents the direction of overall imbalance at each site (compare to **Fig. 3d**).

Nature Genetics: doi:10.1038/ng.3432



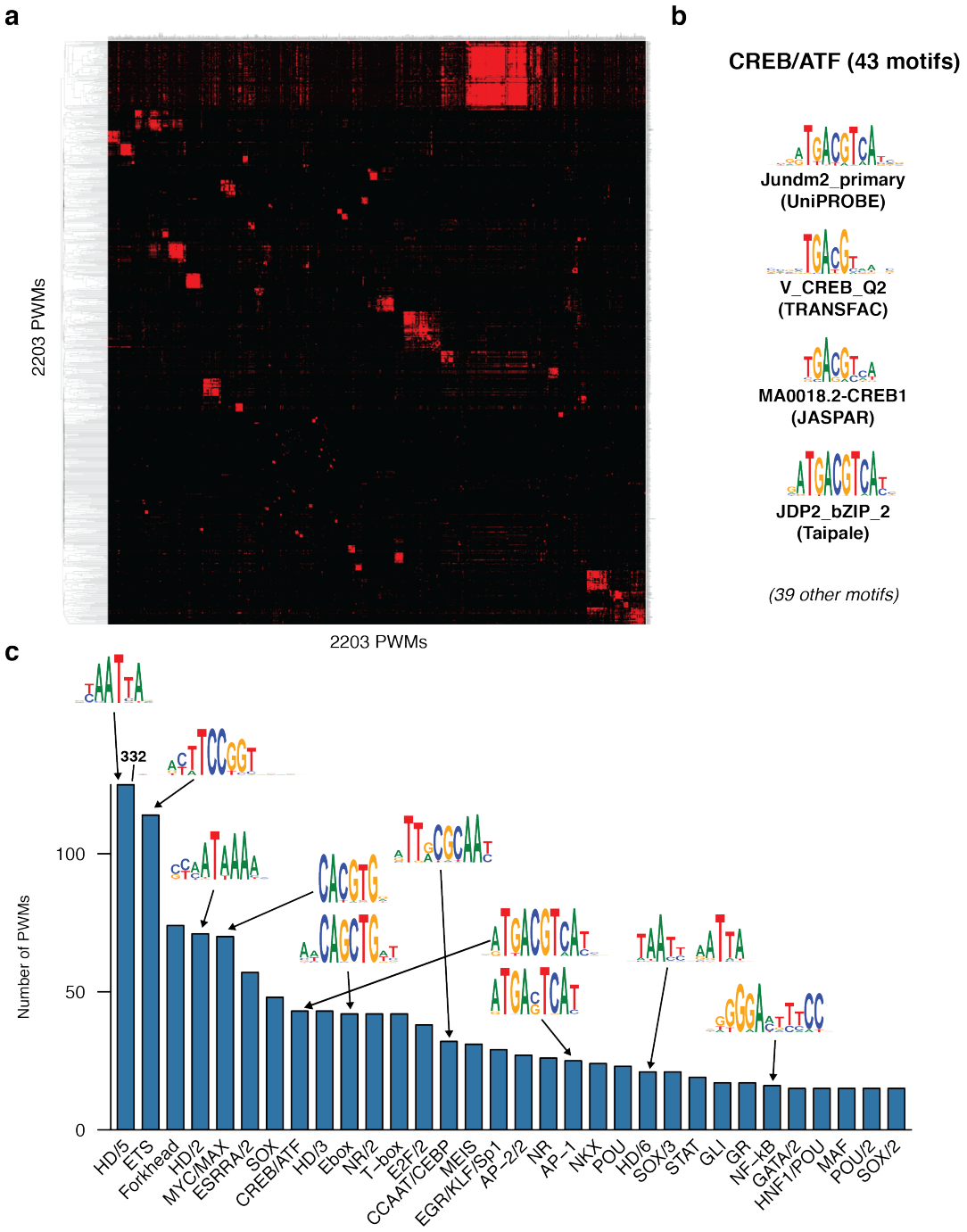
Supplementary Fig. 9. Imbalance in CTCF occupancy and H3K4me3.
(a-b) Strategy and data used to identify imbalanced variants for CTCF (a) and H3K4me3 (b).

(c-d) Allelic ratios of CTCF occupancy (c) and H3K4me3 signal (d)

(e-f) Power to detect imbalance in CTCF occupancy (e) and H3K4me3 (f) increases with total read depth.

(g-h) Negative control showing poor allelic consistency between CTCF (g) and H3K4me3 (h) with nearest non-overlapping DNase I SNP; compare to **Fig. 4c-d**. Shown are sites imbalanced for both. *r*, Pearson correlation of allelic ratios. ChIP-seq and DNase I SNPs were considered imbalanced at 5% FDR.

Nature Genetics: doi:10.1038/ng.3432



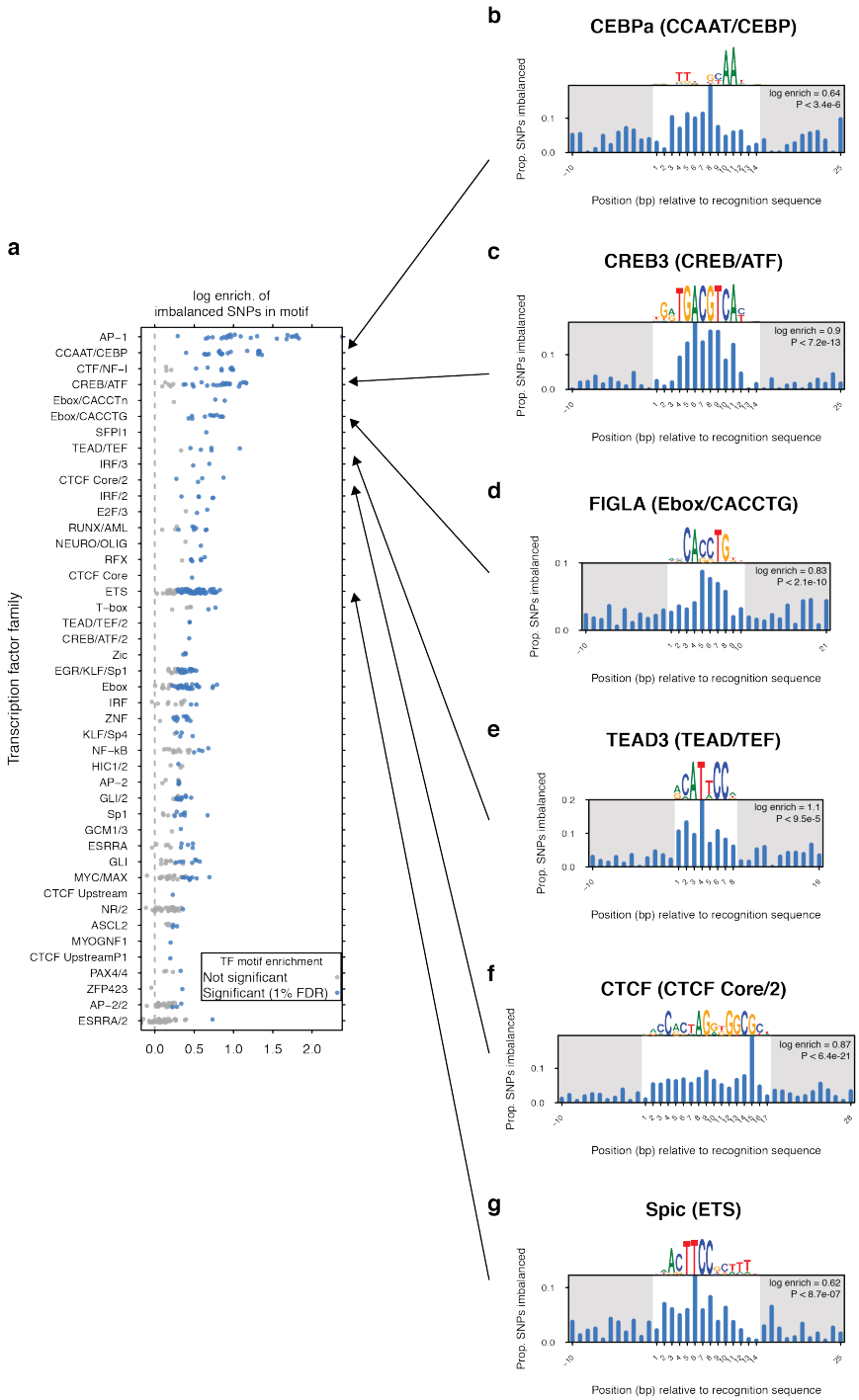
Supplementary Fig. 10. Clustering of motifs into groups of related sequence specificities. Motifs for 2,203 human and mouse sequence-specific factors from the JASPAR, UniProbe, TRANSFAC, and Jolma et al. 2013[35] databases were clustered using similarity scores from the program TOMTOM.

(a) Pairwise similarity scores between motifs. The resulting dendrogram was cut to create nonredundant groups of factors with related sequence specificities.

(b) Example of TF cluster for CREB/ATF, showing related motifs.

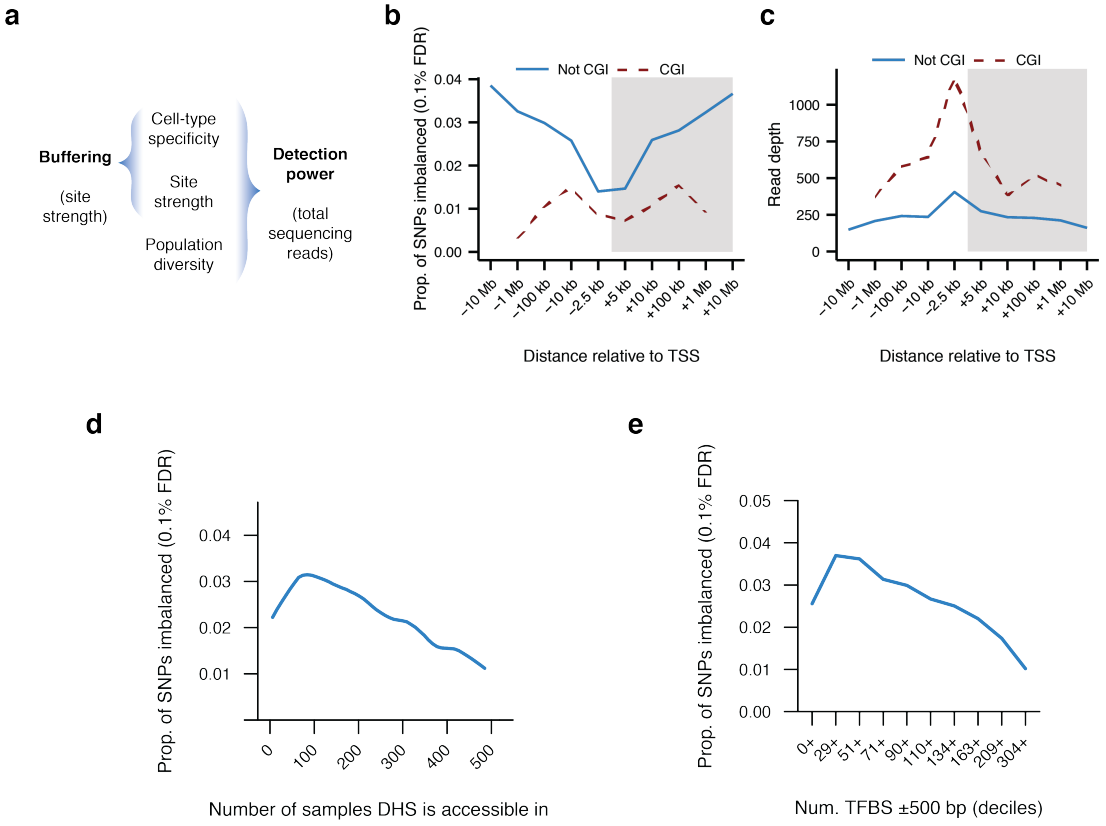
(c) Overview showing number of TF motifs in each cluster.

Nature Genetics: doi:10.1038/ng.3432



Supplementary Fig. 11. Enrichment of imbalanced variants in TF motifs.
Enrichment was computed as the $\log_2$ of the proportion of imbalanced SNPs lying within the recognition sequence divided by the proportion of non-imbalanced SNPs lying within the recognition sequence (a) Detail of **Fig. 5h**. Significance of enrichment of significant SNPs in motifs was assessed by permutation. (b-g) Example motifs showing proportion of imbalanced SNPs per position.

Nature Genetics: doi:10.1038/ng.3432

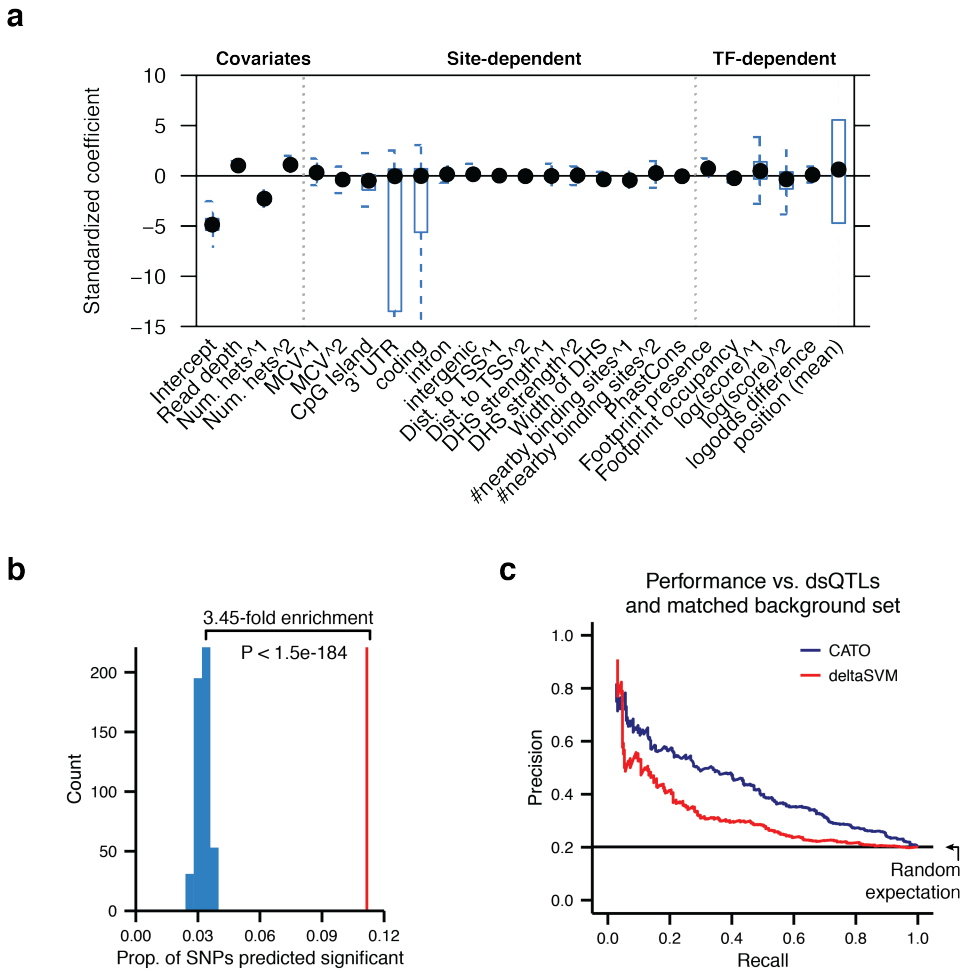


Supplementary Fig. 12. Genomic characteristics associated with site-dependent buffering.
(a) Detection power is determined by total read depth at a site, which is a function of the cell-type selectivity of the DHS, its strength in a single cell type, and the frequency of heterozygous variants in the population. In contrast, population diversity is unrelated to buffering that is a function of site strength.

(b-c) Frequency of (b) imbalance and (c) read depth relative to transcription start sites and broken down by the presence or absence of a CpG island (CGI).

(d-e) Frequency of imbalance relative to broken down by (d) DHS cell-type activity spectrum; or (e) number of nonredundant TF recognition sequences overlapping DNase I footprints nearby.

Nature Genetics: doi:10.1038/ng.3432



Supplementary Fig. 13. TF-centric prediction of variants affecting DNA accessibility.
(a) Summary of standardized regression coefficients across all motifs. Points represent median and boxes indicate interquartile range. Position represents the mean of all coefficients per motif. Features are described in **Online Methods**. See also **Supplementary Table 17**.

(b) Enrichment of dsQTLs[4] in SNPs predicted to affect TF occupancy (score > 0.1).

(c) Precision-recall curve of predictions using dsQTLs as reference set.

Nature Genetics: doi:10.1038/ng.3432

SUPPLEMENTARY TABLES

Supplementary Table 1. Overview of DNase I data used in this study.
DNase I mapping of 114 cell types and tissues used in the study, including the shorthand name for the tissue. Signal Portion of Tags (SPOT) scores are a measure of enrichment and refer to the proportion of reads mapping within DHS. Read counts include reads mapped uniquely with ≤2 mismatches to an autosomal chromosome; paired end reads were required to both properly map to the same chromosome. Read counts are in millions. *, FL_E was excluded from the primary analysis and used for independent validation of the predictions in **Fig. 7**. Previously published data sets are labeled by publication[2,3,24,27,64-67].

Due to the size of this table, it has been made available as a separate tab-delimited text file. The first five rows of the table are shown below:

Sample	DS#	Individual	Cell type	SPOT Score	#Mapped reads	%Dupl. reads	#Nonredundant reads	GEO Accession	Published?
AG04449	DS12319	AG04449_and_AG04450	AG04449	0.47	22.1	14.83%	18.8	GSM736562	Ref. 2
AG04449	DS12329	AG04449_and_AG04450	AG04449	0.39	24.4	12.93%	21.2	GSM736590	Ref. 2
AG04450	DS12255	AG04449_and_AG04450	AG04450	0.41	28.4	13.99%	24.4	GSM736563	Ref. 2
AG04450	DS12270	AG04449_and_AG04450	AG04450	0.45	22.2	15.85%	18.6	GSM736514	Ref. 2

Supplementary Table 2. Overview of ChIP-seq data used in this study.
ChIP-seq mapping of CTCF and H3K4me3 in 49 unique cell types and tissues used in the study, Signal Portion of Tags (SPOT) scores are a measure of enrichment and refer to the proportion of reads mapping within DHS. Read counts include reads mapped uniquely with ≤2 mismatches to an autosomal chromosome; paired end reads were required to both properly map to the same chromosome. Read counts are in millions. Previously published data sets are labeled by publication[2,17,44,68].

Due to the size of this table, it has been made available as a separate tab-delimited text file. The first five rows of the table are shown below:

Sample	DS#	Individual	Cell Type	SPOT Score	#Mapped reads	%Duplicate reads	#Nonredundant reads	GEO Accession	Published?
AG04449	DS15736	AG04449_and_AG04450	AG04449-CTCF	0.27	22.9	8%	21.1	GSE30263	Ref. 68
AG04449	DS15737	AG04449_and_AG04450	AG04449-H3K4me3	0.75	22.7	17%	18.9	GSE35583	Ref. 2
AG04449	DS15738	AG04449_and_AG04450	AG04449-CTCF	0.62	9.7	21%	7.7	GSE30263	Ref. 68
AG04449	DS15739	AG04449_and_AG04450	AG04449-H3K4me3	0.71	23.7	16%	20	GSE35583	Ref. 2

Nature Genetics: doi:10.1038/ng.3432

Supplementary Table 3. Summary of 183 individuals in this study.

All heterozygotes were counted regardless of whether the site was tested for allelic imbalance. Average depth at heterozygous sites includes all DNase I and ChIP-seq reads passing filters. *, FL-E was excluded from the primary analysis and used for independent validation of the predictions in **Fig. 7**.

Individual	# Mapped reads	# Nonredundant reads	# Samples	# Heterozygous sites	Avg. reads per site
AA4	30,503,671	28,617,173	1	851	16
AG04449_and_AG04450	263,596,662	224,592,730	12	26,917	52
AG09309	119,763,712	94,295,357	6	15,832	36
AG09319	475,209,050	417,419,959	8	48,723	64
AG10803	639,578,474	384,936,440	8	41,967	76
AoAF	428,601,109	237,997,463	6	40,359	38
BJ	106,751,332	80,782,256	6	14,558	30
DS14732	23,734,091	22,370,312	1	451	16
DS16376	33,742,948	29,958,187	1	1,058	16
DS17425	27,604,081	23,515,341	1	1,602	21
DS17502	28,419,277	24,290,530	1	1,494	22
FL-E *	733,202,670	559,916,500	2	41,196	68
GM04503_and_GM04504	117,393,212	90,059,669	4	17,279	35
GM06990	187,093,361	106,751,197	5	12,029	33
GM12864	107,256,972	87,694,456	6	12,533	32
GM12865	350,352,823	217,320,097	7	24,051	39
GM12866	41,553,939	34,501,190	3	5,273	21
GM12867	21,647,658	18,570,852	2	1,330	23
GM12868	22,002,502	18,078,615	2	1,231	22
GM12869	18,936,502	15,535,383	2	1,295	22
GM12870	12,054,461	11,471,663	2	967	21
GM12871	23,635,587	18,726,219	2	1,725	26
GM12872	42,894,505	35,031,278	3	2,958	30
GM12873	41,875,027	31,581,809	3	3,506	33
GM12874	21,940,138	16,235,171	2	1,613	25
GM12875	34,473,622	30,150,191	3	3,915	20
GM12878	102,279,829	89,454,692	6	13,252	30
H-22337	59,109,772	51,525,528	2	3,011	20
H-22510	227,588,444	116,216,799	2	20,690	28
H-22662	281,344,601	174,129,857	2	31,540	25
H-22676	49,365,481	41,113,276	2	2,820	24
H-22727	49,836,996	40,742,397	2	5,649	21
H-22911	24,299,300	19,186,043	1	2,063	20
H-22934	42,518,959	33,697,828	1	3,664	20
H-23090	40,560,221	35,515,410	1	2,428	17
H-23247	20,182,753	18,250,821	1	856	17
H-23266	331,372,078	198,557,038	4	35,950	26
H-23284	80,340,490	68,172,018	4	10,243	26
H-23327	32,000,795	30,001,153	1	970	16
H-23365	21,208,419	19,715,705	1	699	16
H-23399	95,726,022	85,868,852	3	8,356	22
H-23419	36,450,602	32,533,165	1	1,463	16
H-23435	115,783,459	104,199,402	4	11,689	24
H-23468	24,757,113	17,039,700	1	101	13
H-23500	48,239,562	42,911,655	2	3,112	18
H-23524	85,160,656	70,320,497	3	8,001	22
H-23548	33,422,500	30,158,954	1	1,775	18
H-23589	131,417,154	119,445,343	4	12,146	22
H-23604	235,161,426	208,300,290	7	33,427	26
H-23617	35,705,323	30,125,180	1	3,001	17
H-23640	207,144,869	157,492,701	5	30,864	31
H-23663	150,914,475	126,035,606	4	16,875	26
H-23724	52,239,272	47,663,291	2	4,145	20
H-23744	157,124,410	134,663,843	5	26,230	28
H-23758	195,754,316	171,522,230	7	26,932	29
H-23769	776,454,341	507,842,619	5	107,187	27
H-23790	53,430,971	46,384,468	2	6,741	23
H-23808	76,329,623	70,347,904	3	6,148	21
H-23833	191,204,880	167,901,301	5	20,863	26
H-23855	64,198,574	59,497,174	2	2,828	18
H-23871	128,956,276	117,931,447	4	13,285	23
H-23887	379,169,242	330,612,241	10	52,610	29
H-23914	619,761,905	369,528,223	8	79,732	29

Nature Genetics: doi:10.1038/ng.3432

Individual	# Mapped reads	# Nonredundant reads	# Samples	# Heterozygous sites	Avg. reads per site
H-23941	185,480,293	156,418,705	5	21,813	28
H-23964	696,989,952	487,646,628	9	114,734	28
H-24005	327,132,379	263,246,912	7	48,071	30
H-24042	154,655,936	114,422,455	3	17,184	26
H-24078	214,802,747	186,555,354	6	25,585	28
H-24089	256,168,350	235,118,611	7	36,998	26
H-24111	348,082,565	319,005,461	10	62,978	27
H-24125	58,626,850	54,259,077	2	3,918	19
H-24143	70,025,911	65,905,115	3	5,577	21
H-24169	101,888,528	94,736,343	4	10,183	24
H-24198	63,887,007	59,662,710	2	4,093	18
H-24218	316,866,434	269,836,793	5	33,420	59
H-24232	110,670,843	100,447,696	4	17,113	25
H-24244	125,129,260	112,076,129	4	14,246	25
H-24259	1,467,118,387	1,201,664,572	19	178,094	71
H-24272	143,951,691	129,274,655	5	17,433	25
H-24279	150,192,798	135,125,962	5	15,088	27
H-24297	128,118,700	111,630,318	4	11,697	25
H-24342	116,600,307	104,163,072	3	11,416	24
H-24365	101,435,127	90,617,933	3	9,137	24
H-24381	24,146,436	21,432,747	1	1,183	18
H-24401	58,413,976	52,564,859	2	3,283	23
H-24409	1,674,088,465	1,145,626,334	7	489,967	27
H-24477	95,217,222	83,173,742	2	9,665	23
H-24493	105,880,520	93,154,906	3	11,979	25
H-24507	91,588,793	82,322,251	3	9,082	24
H-24510	370,454,291	317,678,912	6	66,160	38
H-24568	83,828,760	73,926,733	2	9,927	22
H-24582	87,417,047	77,931,143	3	9,078	23
H-24584	52,899,136	48,407,060	2	3,940	20
H-24595	97,254,502	88,141,121	4	10,914	32
H-24608	68,930,445	60,213,058	2	7,093	22
H-24626	138,525,555	111,020,107	4	20,674	30
H-24639	28,361,326	25,762,120	1	3,861	21
H-24644	56,504,122	50,798,791	2	7,988	32
H-24800	50,598,696	39,852,916	1	6,686	28
H-24851	35,051,810	29,524,576	1	4,635	23
H-24996	24,708,947	23,152,142	1	2,227	18
h.f.Retina-DS23179	33,849,476	28,353,468	1	2,836	20
h.f.Retina-DS23286	45,628,983	33,397,364	1	1,635	16
h.f.Retina-DS23368	42,163,089	29,945,687	1	5,430	21
H1	868,728,355	706,525,426	7	149,679	37
H7	617,533,259	446,510,537	17	70,270	41
HA-h	322,223,186	194,523,503	2	39,584	25
HA-sp	338,968,140	193,663,583	6	30,984	33
HAc_and_HBMEC	286,680,301	231,446,816	12	28,557	47
HAEpiC	287,580,838	128,587,327	2	23,085	31
HBVP	44,035,582	29,151,856	1	1,269	18
HBVSMC	24,976,126	21,830,140	1	1,169	16
HCF	292,571,240	155,482,853	4	21,402	33
HCFaa	378,824,156	310,068,708	4	39,288	58
HCM	322,661,724	180,032,313	6	28,526	39
HConF	70,253,870	54,804,454	2	7,037	27
HCPEpiC	358,985,956	203,840,568	6	34,150	40
HEEpiC	381,367,650	240,945,235	6	38,792	38
HFF	262,504,425	164,437,112	4	31,621	33
HGF	312,674,665	264,279,985	2	19,560	66
HIPEpiC	236,477,404	143,477,982	2	23,843	27
HMEC	123,499,894	102,412,547	6	14,703	34
HMF	627,853,588	353,609,565	8	45,044	82
HMVEC-dAd	24,446,112	21,370,411	1	1,232	18
HMVEC-dBl-Ad	256,490,186	128,486,416	2	20,687	33
HMVEC-dBl-Neo	318,861,813	184,625,792	2	24,560	30
HMVEC-dLy-Ad	66,242,403	47,454,091	2	7,000	30
HMVEC-dLy-Neo	288,398,678	165,067,011	2	19,672	31
HMVEC-dNeo	65,878,969	52,424,183	2	5,505	27
HMVEC-LBl	385,242,594	273,899,866	2	31,423	62
HMVEC-LLy	275,049,373	141,109,352	2	17,861	31
HNPCEpiC	42,034,161	35,866,572	2	4,563	21
HPAEC	31,538,184	28,378,965	1	1,301	17
HPAF	346,249,375	207,379,912	6	29,053	41
HPdLF	285,038,422	147,417,006	2	23,090	28

Nature Genetics: doi:10.1038/ng.3432

Individual	# Mapped reads	# Nonredundant reads	# Samples	# Heterozygous sites	Avg. reads per site
HPF	384,266,980	231,873,791	6	33,838	37
HRCE	234,225,189	133,758,875	2	19,479	27
HRE	96,949,036	73,627,718	6	14,786	29
HRGEC	40,549,464	36,316,223	2	2,961	20
HRPEpiC	141,126,701	115,132,100	6	18,256	34
HSMM	599,457,143	441,224,443	4	57,855	70
hTH-CWilson	148,416,501	77,498,618	1	5,045	21
hTH17-DS11039	178,100,445	109,931,892	1	5,954	48
HUVEC	461,605,192	319,661,722	6	37,540	31
IMR90	296,816,168	196,585,383	4	25,876	33
LHCN_M2	463,961,995	370,740,448	4	46,373	63
NH-A_and_NHLF	529,084,585	273,269,948	4	42,784	38
NHDF_JThompson	319,457,142	298,680,131	4	44,885	36
NHDF-Ad	267,694,855	115,059,171	2	23,443	32
NHDF-neo	449,316,027	241,525,969	6	35,042	38
NHEK	97,783,531	80,114,126	6	10,771	27
NHLF_DIFFERENT	64,754,094	55,968,741	3	9,606	29
PrEC	65,010,554	58,538,577	2	4,555	23
RO_00738	28,857,005	24,286,303	1	1,518	19
RO_01492	48,496,186	35,848,294	2	4,923	26
RO_01507	31,316,639	26,346,614	1	855	20
RO_01508	985,798,063	750,828,704	13	115,074	58
RO_01517	23,868,212	19,590,744	1	1,850	19
RO_01520	32,163,624	23,332,207	1	2,627	21
RO_01525	23,511,152	19,046,813	1	1,564	20
RO_01527	27,143,129	23,090,532	1	1,466	19
RO_01535	233,164,426	115,331,716	2	16,838	29
RO_01536	49,894,735	42,138,272	2	3,706	26
RO_01549	362,142,056	264,751,342	3	32,202	69
RO_01562	32,634,795	26,039,058	1	2,440	20
RO_01679	2,513,490,051	1,774,361,166	12	567,617	41
RO_01689	115,626,075	103,333,547	5	9,155	31
RO_01746	160,255,783	98,765,711	3	12,433	35
RO_01778	583,692,144	395,940,800	4	47,867	58
RO_01794	52,889,678	44,262,428	2	8,020	27
RPTEC	128,139,945	107,947,481	6	21,004	33
SAEC	289,160,907	180,139,393	6	27,864	32
Skin_01	665,225,054	410,063,480	6	77,594	40
Skin_02	193,981,779	154,627,547	6	21,990	33
SkMC	599,813,956	217,933,605	4	48,338	38
STL001	294,235,137	263,132,193	2	40,654	38
STL002	81,971,154	73,970,727	2	4,719	23
STL003	291,943,703	262,174,737	3	29,328	45
vHMEC	220,747,033	195,545,100	2	24,678	50
Wb11970640	264,069,924	215,022,137	1	14,103	61
Wb33676984	600,678,705	489,645,150	3	47,483	50
Wb54553204	693,519,623	531,212,420	4	45,852	79
Wb78495824	197,530,355	153,373,565	1	14,810	59

Nature Genetics: doi:10.1038/ng.3432

Supplementary Table 4. Summary of cell and tissue types defined in this study.
Note that DNase I, CTCF and H3K4me3 samples are listed separately. *, samples were not used in the primary analysis.

Cell type	# Mapped reads	# Nonredundant reads	# Samples
AG04449	46,455,677	40,029,165	2
AG04449-CTCF	32,662,279	28,776,381	2
AG04449-H3K4me3	46,440,531	38,831,689	2
AG04450	50,543,037	43,060,902	2
AG04450-CTCF	37,814,595	34,162,369	2
AG04450-H3K4me3	49,680,543	39,732,224	2
AG09309	51,476,789	40,026,117	2
AG09309-CTCF	23,510,150	21,149,420	2
AG09309-H3K4me3	44,776,773	33,119,820	2
AG09319	381,755,253	343,256,468	4
AG09319-CTCF	46,100,804	37,559,643	2
AG09319-H3K4me3	47,352,993	36,603,848	2
AG10803	247,965,319	115,955,129	2
AG10803-CTCF	54,943,820	47,340,835	2
AG10803-exon *	100,488,810	67,011,274	1
AG10803-H3K4me3	52,997,759	43,344,344	2
AG10803-MS *	183,182,766	111,284,858	1
AoAF	345,753,264	168,855,365	2
AoAF-CTCF	37,514,969	32,507,080	2
AoAF-H3K4me3	45,332,876	36,635,018	2
BJ	50,914,340	37,261,813	2
BJ-CTCF	28,100,072	19,649,270	2
BJ-H3K4me3	27,736,920	23,871,173	2
CD14	524,734,927	389,767,859	5
CD14-H3K4me3	42,949,279	35,870,830	2
CD19_CD20	1,106,932,002	770,542,875	7
CD19-H3K4me3	25,040,913	22,963,038	1
CD20-H3K4me3	59,992,370	50,135,587	3
CD3	369,321,097	226,806,316	4
CD3-H3K4me3	31,006,720	26,651,514	1
CD34	883,269,510	602,480,935	16
CD4	795,391,612	599,301,997	6
CD4_CD8_T-cell	19,284,329	17,107,009	1
CD4-H3K4me3	36,013,269	28,014,310	1
CD4pos_N	264,069,924	215,022,137	1
CD56	561,185,758	424,208,093	3
CD56-H3K4me3	22,040,687	19,992,458	1
CD8	868,431,821	608,544,875	5
CD8-H3K4me3	20,696,488	18,380,083	1
fAdrenal	212,148,751	193,884,223	7
fAdrenal-H3K4me3	50,598,696	39,852,916	1
fBrain	533,969,730	380,857,422	13
fHeart	609,944,127	442,286,129	12
fIntestine	28,890,159	27,068,825	1
fIntestine_Lg	875,174,623	647,597,620	15
fIntestine_Lg-H3K4me3	21,185,122	19,182,951	1
fIntestine_Sm	635,538,822	495,753,533	11
fIntestine_Sm-H3K4me3	27,265,730	24,500,223	1
fKidney	2,103,868,808	1,867,931,045	64
FL-E *	733,202,670	559,916,500	2
fLung	1,510,616,328	1,166,050,281	37
fMuscle	2,656,725,545	1,931,863,134	50
fMuscle_upper_trunk	42,982,637	35,944,094	1
fMuscle-H3K4me3	65,737,576	56,834,270	2
fPlacenta	530,691,296	456,851,655	6
fPlacenta-H3K4me3	24,708,947	23,152,142	1
fSkin	31,486,077	29,382,674	1
fSkin_fibro	1,585,252,475	1,288,519,344	16
fSpinal_cord	482,124,400	357,291,718	5
fSpleen	27,236,227	25,425,282	1
fStomach	706,318,353	543,130,495	12
fStomach-H3K4me3	28,361,326	25,762,120	1
fThymus	647,704,560	533,327,315	11
fThymus-H3K4me3	25,818,356	23,489,097	1
Gastric_Mucosa	76,412,810	71,022,471	2
GM04503_and_GM04504_fibro	117,393,212	90,059,669	4

Nature Genetics: doi:10.1038/ng.3432

Cell type	# Mapped reads	# Nonredundant reads	# Samples
h.f.Retina	121,641,548	91,696,519	3
H1	57,701,013	53,015,377	2
H7	255,869,071	153,821,721	2
H7_hESC_T14	53,669,162	46,117,617	2
H7_hESC_T14-H3K4me3	38,344,714	27,780,664	2
H7_hESC_T2	23,734,091	22,370,312	1
H7_hESC_T2-H3K4me3	39,545,740	32,105,895	2
H7_hESC_T5	68,368,180	61,262,122	2
H7_hESC_T5-H3K4me3	31,026,750	22,145,157	2
H7_hESC_T9	67,847,819	54,161,745	1
H7_hESC_T9-H3K4me3	33,402,320	28,364,621	2
H7-H3K4me3	29,459,503	20,750,995	2
HA-h	322,223,186	194,523,503	2
HA-sp	267,416,882	133,803,490	2
HA-sp-CTCF	27,634,276	25,097,647	2
HA-sp-H3K4me3	43,916,982	34,762,446	2
HAc	67,277,958	55,972,552	2
HAc-CTCF	36,243,237	29,605,435	2
HAc-H3K4me3	36,997,962	28,228,464	2
HAEpiC	287,580,838	128,587,327	2
HBMEC	66,617,037	55,396,270	2
HBMEC-CTCF	37,626,662	30,065,560	2
HBMEC-H3K4me3	41,917,445	32,178,535	2
HBVP	44,035,582	29,151,856	1
HBVSMC	24,976,126	21,830,140	1
HCF	252,370,999	122,693,384	2
HCF-H3K4me3	40,200,241	32,789,469	2
HCFaa	332,628,365	274,762,199	2
HCFaa-CTCF	19,720,630	16,377,267	1
HCFaa-H3K4me3	26,475,161	18,929,242	1
HCM	246,163,743	119,729,865	2
HCM-CTCF	31,107,056	26,859,610	2
HCM-H3K4me3	45,390,925	33,442,838	2
HConF	70,253,870	54,804,454	2
HCPEpiC	260,018,096	126,322,669	2
HCPEpiC-CTCF	51,197,777	40,471,266	2
HCPEpiC-H3K4me3	47,770,083	37,046,633	2
Heart	30,503,671	28,617,173	1
HEEpiC	288,251,229	163,260,558	2
HEEpiC-CTCF	32,806,611	28,939,753	2
HEEpiC-H3K4me3	60,309,810	48,744,924	2
HFF	230,270,593	139,087,383	2
HFF-CTCF	13,923,171	10,553,042	1
HFF-H3K4me3	18,310,661	14,796,687	1
HGF	312,674,665	264,279,985	2
HIPEpiC	236,477,404	143,477,982	2
HMEC	54,521,984	44,866,676	2
HMEC-CTCF	31,440,642	26,690,161	2
HMEC-H3K4me3	37,537,268	30,855,710	2
HMF	289,675,923	129,569,099	2
HMF-CTCF	29,942,543	26,667,169	2
HMF-exon *	85,888,176	55,301,472	1
HMF-H3K4me3	50,466,601	40,790,827	2
HMF-MS *	171,880,345	101,280,998	1
HMVEC-dAd	24,446,112	21,370,411	1
HMVEC-dBl-Ad	256,490,186	128,486,416	2
HMVEC-dBl-Neo	318,861,813	184,625,792	2
HMVEC-dLy-Ad	66,242,403	47,454,091	2
HMVEC-dLy-Neo	288,398,678	165,067,011	2
HMVEC-dNeo	65,878,969	52,424,183	2
HMVEC-LBl	385,242,594	273,899,866	2
HMVEC-LLy	275,049,373	141,109,352	2
HNPCEpiC	42,034,161	35,866,572	2
HPAEC	31,538,184	28,378,965	1
HPAF	242,385,834	124,191,322	2
HPAF-CTCF	53,319,350	44,939,051	2
HPAF-H3K4me3	50,544,191	38,249,539	2
HPdLF	285,038,422	147,417,006	2
HPF	298,755,480	157,637,887	2
HPF-CTCF	44,094,150	40,427,700	2
HPF-H3K4me3	41,417,350	33,808,204	2
HRCE	234,225,189	133,758,875	2

Nature Genetics: doi:10.1038/ng.3432

Cell type	# Mapped reads	# Nonredundant reads	# Samples
HRE	45,756,248	38,991,432	2
HRE-CTCF	24,473,735	11,818,796	2
HRE-H3K4me3	26,719,053	22,817,490	2
HRGEC	40,549,464	36,316,223	2
HRPEpiC	51,161,485	39,522,715	2
HRPEpiC-CTCF	39,037,569	34,793,162	2
HRPEpiC-H3K4me3	50,927,647	40,816,223	2
HSMM	546,086,436	395,588,101	2
HSMM_D	53,370,707	45,636,342	2
hTH1	465,066,570	348,602,687	4
hTH17	178,100,445	109,931,892	1
hTH2	291,039,735	204,019,991	2
hTreg	197,530,355	153,373,565	1
HUVEC	410,864,741	282,045,137	2
HUVEC-CTCF	18,992,505	10,499,100	2
HUVEC-H3K4me3	31,747,946	27,117,485	2
IMR90	296,816,168	196,585,383	4
iPS_JThompson	319,457,142	298,680,131	4
iTH1	329,814,792	259,176,241	1
iTH2	356,693,732	286,557,269	1
LCL	451,068,633	266,883,370	6
LCL-CTCF	379,000,590	292,031,306	32
LCL-H3K4me3	197,927,703	172,188,137	10
LHCN_M2	216,005,079	173,682,243	2
LHCN_M2_D4	247,956,916	197,058,205	2
Mesendoderm	277,079,640	186,535,812	2
MSC_mesenchymal	325,536,875	270,792,026	2
NH-A	216,209,064	128,944,812	2
NHDF-Ad	267,694,855	115,059,171	2
NHDF-neo	349,669,504	156,972,405	2
NHDF-neo-CTCF	38,721,897	34,771,803	2
NHDF-neo-H3K4me3	60,924,626	49,781,761	2
NHEK	54,968,748	49,231,516	2
NHEK-CTCF	20,675,601	11,794,943	2
NHEK-H3K4me3	22,139,182	19,087,667	2
NHLF	312,875,521	144,325,136	2
NHLF-CTCF	24,019,094	20,924,772	1
NHLF-H3K4me3	40,735,000	35,043,969	2
Ovary	40,714,667	36,662,029	1
Pancreas	75,717,014	68,461,122	2
PrEC	65,010,554	58,538,577	2
Psoas_Muscle	255,119,911	226,840,353	1
RPTEC	49,630,395	42,427,440	2
RPTEC-CTCF	37,181,692	31,105,119	2
RPTEC-H3K4me3	41,327,858	34,414,922	2
SAEC	240,423,285	144,627,295	2
SAEC-CTCF	22,318,664	12,780,068	2
SAEC-H3K4me3	26,418,958	22,732,030	2
Skin_Fibroblasts	469,276,754	226,113,365	4
Skin_Keratinocytes	241,908,438	212,858,666	4
Skin_Melanocytes	148,021,641	125,718,996	4
SkMC	556,559,495	185,101,894	2
SkMC-H3K4me3	43,254,461	32,831,711	2
Small_Bowel_Mucosa	220,185,592	196,291,682	1
Trophoblast	208,410,827	196,182,211	1
vHMEC	220,747,033	195,545,100	2

Nature Genetics: doi:10.1038/ng.3432

Supplementary Table 5. Genotyping sensitivity and specificity relative to HAIB data.

Individual ID	HAIB sample	# HAIB hets in DHS	# HAIB hets passing our genotyping filters	# hets in DHS	Concordant hets	HAIB Sensitivity	Raw sensitivity	Sensitivity of filtered genotypes	# FP hets	FPR
AG04449_and_AG04450	AG04449_rep1	14,207	4,440	18,437	4,420	24.0%	31.1%	99.5%	11	0.25%
AG04449_and_AG04450	AG04450_rep2	14,187	4,434	18,437	4,414	23.9%	31.1%	99.5%	17	0.38%
AG09309	AG09309_rep1	10,212	2,536	9,366	2,522	26.9%	24.7%	99.4%	10	0.39%
AG09319	AG09319_rep2	10,116	5,256	19,300	5,237	27.1%	51.8%	99.6%	13	0.25%
AG10803	AG10803_rep1	11,045	5,192	19,116	5,162	27.0%	46.7%	99.4%	18	0.35%
BJ	BJ_rep1	10,539	2,476	9,200	2,459	26.7%	23.3%	99.3%	10	0.41%
GM06990	GM06990_rep1	8,011	1,767	6,924	1,749	25.3%	21.8%	99.0%	8	0.46%
GM12878	GM12878_rep1	9,206	2,160	7,607	2,141	28.1%	23.3%	99.1%	8	0.37%
H1	H1.hES_rep2	13,417	8,673	34,631	8,641	25.0%	64.4%	99.6%	48	0.55%
HAEpiC	HAEpiC_rep1	5,845	3,025	10,972	2,973	27.1%	50.9%	98.3%	11	0.37%
HCF	HCF_rep1	5,674	2,792	10,522	2,780	26.4%	49.0%	99.6%	11	0.39%
HCF	HCF_rep2	5,661	2,782	10,522	2,770	26.3%	48.9%	99.6%	21	0.75%
HCM	HCM_rep1	10,521	4,291	16,330	4,272	26.2%	40.6%	99.6%	22	0.51%
HCPEpiC	HCPEpiC_rep1	13,215	5,397	20,459	5,358	26.2%	40.5%	99.3%	10	0.19%
HCPEpiC	HCPEpiC_rep2	13,211	5,394	20,459	5,355	26.2%	40.5%	99.3%	13	0.24%
HIPEpiC	HIPEpiC_rep1	7,140	3,325	12,123	3,313	27.3%	46.4%	99.6%	16	0.48%
HIPEpiC	HIPEpiC_rep2	7,150	3,334	12,123	3,322	27.4%	46.5%	99.6%	7	0.21%
HMEC	HMEC_rep2	10,402	2,491	9,681	2,479	25.6%	23.8%	99.5%	13	0.52%
HRCE	HRCE_rep1	5,730	2,699	9,711	2,684	27.6%	46.8%	99.4%	4	0.15%
HRCE	HRCE_rep2	5,736	2,701	9,711	2,686	27.7%	46.8%	99.4%	2	0.07%
HRE	HRE_rep1	11,005	2,167	9,254	2,155	23.3%	19.6%	99.4%	9	0.42%
HRE	HRE_rep2	11,011	2,165	9,254	2,153	23.3%	19.6%	99.4%	11	0.51%
HRPEpiC	HRPE_rep2	11,746	2,965	10,860	2,953	27.2%	25.1%	99.6%	3	0.10%
IMR90	IMR90_rep1	7,564	3,538	14,084	3,499	24.8%	46.3%	98.9%	4	0.11%
IMR90	IMR90_rep2	7,566	3,536	14,084	3,497	24.8%	46.2%	98.9%	6	0.17%
NH-A_and_NHLF	NHA_rep1	6,932	4,245	16,188	4,214	26.0%	60.8%	99.3%	25	0.59%
NH-A_and_NHLF	NHA_rep2	6,941	4,256	16,188	4,225	26.1%	60.9%	99.3%	14	0.33%
NHDF-neo	NHDF_neo_rep1	10,525	4,820	17,603	4,802	27.3%	45.6%	99.6%	7	0.15%
RPTEC	RPTEC_rep1	12,249	3,195	13,068	3,190	24.4%	26.0%	99.8%	6	0.19%
RPTEC	RPTEC_rep2	12,246	3,193	13,068	3,188	24.4%	26.0%	99.8%	8	0.25%
SAEC	SAEC_rep1	10,473	4,180	16,000	4,159	26.0%	39.7%	99.5%	10	0.24%
SAEC	SAEC_rep2	10,479	4,178	16,000	4,156	26.0%	39.7%	99.5%	13	0.31%
SkMC	SKMC_rep1	7,970	4,732	19,332	4,683	24.2%	58.8%	99.0%	8	0.17%
SkMC	SKMC_rep2	7,961	4,731	19,332	4,682	24.2%	58.8%	99.0%	9	0.19%
Average:						25.8%	40.4%	99.4%		0.30%

Comparison of heterozygotes identified, sensitivity and specificity of genotype calls from this work and published HAIB Illumina arrays. FP, number of false positive heterozygote calls. FPR, rate of false positive heterozygote calls. HAIB sensitivity was computed relative to all SNPs in DHSs (1% FDR hotspot peaks). FP, FPR, raw sensitivity (**Supplementary Fig. 2a**) were computed considering all HAIB SNPs in DHSs and on the Illumina design. Sensitivity of filtered genotypes was calculated considering only sites where we had a genotype call passing all filters (**Supplementary Fig. 2b**).

While raw sensitivity (30-70%) reflects the performance of our approach relative to independent genotyping, it is penalized by our stringent filtering. In contrast, sensitivity at genotypes passing these filters (99.4%) assesses the potential for artificial imbalance stemming from genotyping error.

Nature Genetics: doi:10.1038/ng.3432

Supplementary Table 6. Rate of imbalance using genotypes confirmed by HAIB Illumina arrays.

Variant set	# SNPs	# Imbalanced SNPs (0.1% FDR)	% Imbalanced SNPs (0.1% FDR)	# Imbalanced SNPs (5% FDR)	% Imbalanced SNPs (5% FDR)
Discovered	14,920	363	2.4%	3,145	21.1%
Discovered & HAIB	14,443	348	2.4%	2,989	20.7%
Discovered & not HAIB	546	16	2.9%	89	16.3%

For the "Discovered" set, imbalance was computed using the subset of samples and discovered SNPs present in the HAIB Illumina genotypes. For the other two sets, imbalance was computed considering only heterozygous samples whose genotypes were confirmed or contradicted, respectively, by the HAIB data. Analysis of imbalance across all heterozygous samples was performed as before, including restriction to sites with >50 reads.

Supplementary Table 7. Location of SNPs tested for imbalance relative to genes.

Genic Location	SNPs tested for imbalance	Imbalanced SNPs (0.1% FDR)	% Imbalanced SNPs (0.1% FDR)	Imbalanced SNPs (5% FDR)	% Imbalanced SNPs (5% FDR)
intergenic	164,189	4,910	3.0%	30,831	18.8%
intron	134,570	3,768	2.8%	24,052	17.9%
promoter	53,516	599	1.1%	8,302	15.5%
coding	6,030	90	1.5%	796	13.2%
3'UTR	3,979	89	2.2%	616	15.5%

Supplementary Table 8. Reference bias affecting targeted footprinting sites.

SNP location relative to probe	Probe melting temperature (°C)	Avg. allelic ratio	#SNPs
Off target	NA	49.63%	3,681
Flanking 250bp	NA	50.05%	1,199
Over probe	[85.9,105)	53.61%	1,068
Over probe	[105,110)	53.07%	1,067
Over probe	[110,113)	52.14%	1,068
Over probe	[113,116)	52.77%	1,068
Over probe	[116,126]	52.00%	1,068

Shown are allelic ratios (the proportion of reads mapping to the reference allele) for targeted data. Allelic ratios are near the expected 50%, except when directly overlapping probes where increased hybridization energy results in correctable reference bias (**Online Methods**).

Nature Genetics: doi:10.1038/ng.3432

Supplementary Table 9. Samples used for correlation of allelic ratios in same individual or cell type.

Data from the same individual and cell type were pooled and used for the analysis in **Fig. 3a**. Samples are labeled as *Cell type*:*Individual*.

Sample	# Heterozygous sites	# Imbalanced sites	Avg. reads per site	# Nonredundant reads (M)
CD14:RO_01679	4879	928	161	270
CD14:RO_01746	1927	201	54	30.5
CD19_CD20:RO_01679	7105	495	118	313.8
CD19_CD20:RO_01778	5986	712	146	118.6
CD34:RO_01492	892	37	44	19.3
CD34:RO_01535	3217	287	48	94.9
CD34:RO_01549	5705	800	131	209.3
CD3:RO_01679	4953	425	51	129.6
CD4:RO_01508	6421	878	170	22.2
CD4:RO_01679	6696	729	116	31.8
CD56:RO_01679	8120	1165	141	367.9
CD8:RO_01508	5122	749	160	214.9
CD8:RO_01679	8550	1383	178	20.7
fIntestine_Lg:H-23769	3289	171	40	222.8
fIntestine_Sm:H-23769	2954	109	38	206.1
fKidney:H-23640	917	41	43	32.5
fKidney:H-24510	4669	340	82	179.3
fKidney:H-24568	928	34	42	32.4
fLung:H-23266	3498	311	46	135
fLung:H-23640	2211	134	46	53.9
fLung:H-23744	1349	60	42	27.7
fLung:H-23758	1400	60	42	25.2
fLung:H-23887	951	94	43	32
fLung:H-23964	978	37	41	51.5
fLung:H-24005	1377	51	43	27.8
fLung:H-24111	934	42	39	33.9
fLung:H-24626	1652	124	48	28
fMuscle:H-23914	6214	315	51	167.7
fMuscle:H-23941	963	38	45	28.2
fMuscle:H-24005	3060	170	46	63.8
fMuscle:H-24042	1477	70	44	48.2
fMuscle:H-24078	984	32	43	21.3
fMuscle:H-24089	1173	45	41	27.9
fMuscle:H-24111	1278	21	40	30
fMuscle:H-24244	1575	82	45	26.4
fMuscle:H-24272	1369	65	43	27.1
fMuscle:H-24279	1211	51	44	17
fMuscle:H-24409	9519	925	64	214.1
fPlacenta:H-24409	5483	472	78	279.7
fSkin_fibro:H-24218	5558	963	120	20.9
fSkin_fibro:H-24259	11437	2305	350	177.6
fSpinal_cord:H-24409	3344	129	39	239.3
fStomach:H-23964	4051	217	41	213.6
fThymus:H-23964	1026	48	44	25.2
fThymus:H-24409	1876	83	39	169.4
HAc:HAc_and_HBMEC	963	64	41	23.6
HBMEC:HAc_and_HBMEC	1155	59	41	28.2
hTH1:Wb33676984	3309	374	125	205.2
hTH2:Wb54553204	4208	817	170	178.8
iTH1:Wb33676984	3070	487	140	259.2
iTH2:Wb54553204	5460	532	128	286.6
LHCN_M2_D4:LHCN_M2	6104	872	88	170.4
LHCN_M2:LHCN_M2	5113	875	85	27.5
Mesendoderm:H1	3521	215	44	21.1
MSC_mesenchymal:H1	7300	1063	117	245.3
NH-A:NH-A_and_NHLF	3839	356	46	110.3
NHLF:NH-A_and_NHLF	5393	661	49	122.3
Skin_Fibroblasts:Skin_01	6933	861	55	145.4
Skin_Fibroblasts:Skin_02	2039	136	47	22.7
Skin_Keratinocytes:Skin_01	4002	378	73	135.5
Skin_Keratinocytes:Skin_02	866	39	42	23.3
Skin_Melanocytes:Skin_01	1280	83	44	31.6
Trophoblast:H1	2868	231	72	196.2

Nature Genetics: doi:10.1038/ng.3432

Supplementary Table 10. Cell types used for identification of context-independent and -dependent sites.

Data from the same cell type were pooled and used for the analysis in **Fig. 3b-e** and **Supplementary Fig. 8**.

Cell type	# Heterozygous sites	# Imbalanced sites	Avg. reads per site	# Nonredundant reads (M)
AG09319	2773	234	64	343.3
AG10803	821	138	59	115.9
AoAF	1148	176	59	168.8
CD14	1717	279	69	389.8
CD19_CD20	3624	252	70	770.5
CD3	640	71	64	226.8
CD34	2825	300	69	602.5
CD4	3634	382	73	599.2
CD4pos_N	810	94	73	215
CD56	2287	288	72	424.2
CD8	4038	479	72	608.6
fBrain	844	50	64	381
fHeart	1407	87	63	442.2
fIntestine_Lg	2043	130	63	647.5
fIntestine_Sm	1545	129	64	495.7
fKidney	7931	293	67	1867.9
fLung	7193	521	66	1166.1
fMuscle	10205	617	66	1931.8
fPlacenta	1728	143	64	456.9
fSkin_fibro	7180	1090	65	1288.3
fSpinal_cord	369	15	57	357.2
fStomach	1142	58	61	543.1
fThymus	887	60	63	533.5
GM04503_and_GM04504_fibro	978	148	62	90.1
H7	290	45	57	153.8
HAEpiC	842	131	59	128.5
HA-h	481	31	55	194.5
HCF	522	50	59	122.7
HCFaa	2527	359	64	274.7
HCM	568	100	58	119.7
HCPEpiC	712	98	58	126.4
HEEpiC	815	98	57	163.3
HFF	677	86	58	139.1
HGF	1492	290	66	264.3
HIPEpiC	587	69	57	143.4
HMF	1149	168	60	129.6
HMVEC-dBl-Ad	849	102	60	128.4
HMVEC-dBl-Neo	807	98	59	184.6
HMVEC-dLy-Neo	707	94	60	165.1
HMVEC-LBl	1798	262	64	273.9
HMVEC-LLy	536	80	59	141.1
HPAF	588	80	59	124.2
HPdLF	651	97	58	147.4
HPF	877	142	60	157.6
HRCE	460	41	57	133.8
HSMM	3632	516	65	395.6
hTH1	999	113	74	348.6
hTH17	282	29	77	109.9
hTH2	1195	220	76	204
hTreg	799	84	74	153.4
HUVEC	398	29	57	282
IMR90	1174	164	60	196.6
iPS_JThompson	1041	62	64	298.7
iTH1	790	144	80	259.2
iTH2	1563	128	72	286.6
LCL	660	94	63	267
LHCN_M2	1586	266	67	173.6
LHCN_M2_D4	1809	239	68	197.1
Mesendoderm	332	37	57	186.5
MSC_mesenchymal	2322	326	64	270.8
NH-A	452	57	58	129
NHDF-Ad	1133	167	59	115.1
NHDF-neo	1001	154	59	156.9
NHLF	913	117	57	144.4
Psoas_Muscle	1106	120	65	226.8
SAEC	467	33	55	144.7

Nature Genetics: doi:10.1038/ng.3432

Skin_Fibroblasts	2067	278	60	226.1
Skin_Keratinocytes	1337	122	62	212.8
Skin_Melanocytes	201	10	59	125.8
SkMC	2136	341	59	185.1
Small_Bowel_Mucosa	1001	100	66	196.3
Trophoblast	637	50	63	196.2
vHMEC	1415	196	64	195.5

Nature Genetics: doi:10.1038/ng.3432

Supplementary Table 11. Co-occurrence of allelic imbalance in DNase I and H3K4me3/CTCF.

		DNase I	
		Imbalanced (5% FDR)	Not imbalanced
CTCF	Imbalanced (5% FDR)	566	244
	Not imbalanced	2,513	8,032
H3K4me3	Imbalanced (5% FDR)	521	713
	Not imbalanced	5,715	24,286

Shown are counts of variants with sufficient power to test for imbalance (**Online Methods**) broken down by the presence or absence of allelic imbalance.

Nature Genetics: doi:10.1038/ng.3432

Supplementary Table 12. Genome-wide counts of TF recognition sequences by motif database.

Motif DB	Num. instances
CTCF_upstream	8,720,005
JASPAR	118,570,999
Taipale SELEX	517,843,584
UniProbe	252,019,007
TRANSFAC	706,858,355

Shown are the total number of occurrences of all motifs in the listed databases (FIMO $P < 10^{-4}$; see **Online Methods**). CTCF_upstream includes 3 models for the extended binding mode of the genomic regulator CTCF[17].

Supplementary Table 13. Coverage of known TF genes and TF clusters by motif database.

Motif DB	# Motifs	# Genes	# Clusters
All	2,203	825	270
CTCF_upstream	3	1	3
JASPAR	145	129	74
Taipale SELEX	843	463	114
TRANSFAC	834	502	191
UniProbe	378	262	99
All (without JASPAR)	2,058	810	265
All (without Taipale SELEX)	1,360	610	234
All (without TRANSFAC)	1,369	620	194
All (without UniProbe)	1,825	786	245

Shown are the number of motifs, unique genes, and TFs clusters represented by each database. The completeness of each databases portrayed by counts excluding that database; note that no single database alone provides a comprehensive set of models.

Supplementary Table 14. Clustering of motifs into TF families.
Clustering of motifs from JASPAR, UniProbe, TRANSFAC, and Jolma et al. 2013[35] databases. Each TF cluster is listed along with the names of constituent motifs.

Due to the size of this table, it has been made available as a separate tab-delimited text file. The first five rows of the table are shown below:

Cluster	Motifs
1.P53	V_P53_05
2.RORA1	V_RORA1_01
3.PR	V_PR_02
4.PAX5	V_PAX5_01

Nature Genetics: doi:10.1038/ng.3432

Supplementary Table 15. TF sensitivity profiles to sequence perturbation.

	Motifs	TF genes	TF clusters
Number with FIMO matches overlapping a SNP	2,148	695	270
Number with sufficient SNPs for analysis	764	350	144
Number with enrichment of imbalanced SNPs in motif (1% FDR)	313	194	44

Sensitivity profiles were generated for each motif with sufficient SNPs overlapping instances genome-wide. Total coverage was estimated by counting the number of genes or TF clusters with at least one motif matching the listed criteria.

Nature Genetics: doi:10.1038/ng.3432

Supplementary Table 16. Summary of imbalanced variants at TF recognition sequences.
#SNPs refers to the number of SNPs within ±20 bp of the TF recognition sequence. The proportion of substitutions imbalanced was computed as the median proportion of SNPs imbalanced at all positions within the TF recognition sequence. Enrichment was computed as $\log_2$ of the proportion of imbalanced SNPs lying within the recognition sequence divided by the proportion of non-imbalanced SNPs lying within the recognition sequence.

Motif	TF Cluster	Motif length (bp)	#SNPs	#Imbalanced SNPs	#SNPs in motif	#Imbalanced SNPs in Motif	Prop. substitutions in motif imbalanced	Enrich. imbalanced SNPs in motif	Q-value
CTCF_UpstreamP1	12-CTCF_UpstreamP1	35	61,852	1,323	31,928	780	0.021	0.2	5.0e-06
V_CTCF_01	16-CTCF_Core	20	49,196	1,171	19,440	636	0.03	0.47	3.3e-19
V_MYOGNF1_01	20.MYOGNF1	29	18,079	452	9,125	261	0.024	0.2	6.7e-03
Ehf	25-ETS	17	9,494	327	3,486	164	0.026	0.47	3.0e-05
EHF_ETS_1	25-ETS	12	6,860	241	2,263	116	0.026	0.57	7.0e-05
Ehf_primary	25-ETS	15	8,007	272	2,883	134	0.015	0.47	1.5e-04
ELF1_ETS_1	25-ETS	12	6,369	227	2,097	112	0.028	0.61	7.5e-05
Elf2	25-ETS	16	7,466	269	2,762	143	0.024	0.55	1.9e-05
Elf3	25-ETS	17	8,083	281	2,986	136	0.023	0.41	8.8e-04
ELF3_ETS_1	25-ETS	12	7,103	257	2,281	120	0.033	0.56	9.0e-05
ELF3_ETS_2	25-ETS	13	7,135	248	2,392	121	0.031	0.57	5.7e-05
Elf3_primary	25-ETS	13	7,167	264	2,114	107	0.032	0.48	1.0e-03
Elf4	25-ETS	16	7,651	281	2,799	151	0.028	0.58	1.5e-06
ELF4_ETS_1	25-ETS	12	7,027	245	2,340	117	0.023	0.54	1.5e-04
Elf5	25-ETS	14	7,773	296	2,404	136	0.039	0.6	2.1e-05
ELF5_ETS_1	25-ETS	11	8,078	274	2,401	117	0.034	0.54	1.4e-04
ELF5_ETS_2	25-ETS	11	8,531	275	2,533	117	0.028	0.54	2.0e-04
Elf5.mouse_ETS_1	25-ETS	11	7,906	261	2,361	114	0.03	0.57	1.2e-04
Elk1	25-ETS	17	7,515	266	2,933	133	0.019	0.37	2.5e-03
ELK1_ETS_4	25-ETS	10	4,236	151	1,488	87	0.034	0.75	1.3e-05
Elk3	25-ETS	17	8,936	283	3,475	138	0.02	0.34	4.5e-03
ELK3_ETS_1	25-ETS	10	4,205	161	1,488	93	0.035	0.74	2.7e-05
Elk4	25-ETS	16	6,290	205	2,458	115	0.02	0.54	2.6e-04
ELK4_ETS_1	25-ETS	10	4,987	193	1,652	105	0.042	0.75	2.0e-06
ERF_ETS_1	25-ETS	10	5,370	223	1,632	100	0.036	0.59	2.3e-04
Erg	25-ETS	16	8,175	317	2,918	164	0.038	0.56	3.3e-06
ERG_ETS_1	25-ETS	10	4,940	210	1,586	108	0.056	0.72	3.7e-06
ERG_ETS_3	25-ETS	10	5,727	239	1,793	120	0.056	0.72	3.6e-06
Ets1	25-ETS	17	7,742	321	2,897	159	0.043	0.42	2.2e-04
ETS1_ETS_1	25-ETS	10	5,632	233	1,759	122	0.045	0.79	2.2e-07
ETS1_ETS_3	25-ETS	10	4,649	192	1,576	99	0.051	0.64	6.6e-05
Etv1	25-ETS	17	8,063	304	3,126	159	0.036	0.45	1.4e-04
ETV2_ETS_1	25-ETS	11	5,542	231	1,696	109	0.052	0.66	2.9e-05
ETV3_ETS_1	25-ETS	10	4,127	170	1,491	88	0.04	0.55	6.4e-04
Etv4	25-ETS	16	6,940	217	2,757	118	0.016	0.47	2.7e-04
Etv5	25-ETS	16	7,001	261	2,720	132	0.033	0.4	1.4e-03
Etv6	25-ETS	17	6,742	234	2,628	132	0.034	0.56	7.7e-06
ETV6_ETS_1	25-ETS	15	13,123	334	4,680	150	0.02	0.34	2.5e-03
ETV6_ETS_2	25-ETS	10	7,895	266	2,327	118	0.036	0.62	3.5e-05
Fli1	25-ETS	16	7,900	301	2,793	154	0.038	0.56	1.3e-05
FLI1_ETS_1	25-ETS	10	5,322	218	1,744	108	0.034	0.63	3.8e-05
FLI1_ETS_3	25-ETS	10	4,898	195	1,608	106	0.051	0.77	6.0e-06
Gabpa	25-ETS	17	5,604	226	2,230	127	0.029	0.52	1.1e-04
GABPA_ETS_1	25-ETS	10	4,823	172	1,715	91	0	0.6	2.6e-04
Gabpa_primary	25-ETS	17	6,156	246	2,394	136	0.026	0.53	3.8e-05
MA0062.2-GABPA	25-ETS	11	9,613	331	2,971	154	0.026	0.62	1.4e-06
MA0136.1-ELF5	25-ETS	9	6,593	248	1,471	80	0.025	0.56	3.0e-03
MA0156.1-FEV	25-ETS	8	7,242	261	1,640	94	0.039	0.7	1.4e-04
Sfpi1	25-ETS	16	10,721	313	3,687	134	0.027	0.33	7.7e-03
Sfpi1_primary	25-ETS	14	10,992	351	3,354	145	0.026	0.45	4.9e-04
SPDEF_ETS_4	25-ETS	11	3,867	131	1,410	72	0.031	0.62	6.9e-04
SPI1_ETS_1	25-ETS	14	8,163	304	2,637	129	0.041	0.41	1.9e-03
SPIB_ETS_1	25-ETS	14	9,149	325	2,876	134	0.033	0.41	1.7e-03
Spic	25-ETS	14	8,284	277	2,653	134	0.031	0.62	6.0e-06
Spic.mouse_ETS_1	25-ETS	14	8,216	307	2,588	127	0.038	0.41	1.7e-03
V_CETS1P54_01	25-ETS	10	8,777	325	2,343	122	0.048	0.51	3.6e-04
V_CETS1P54_02	25-ETS	13	6,493	240	2,154	107	0.036	0.45	2.2e-03
V_CETS1P54_03	25-ETS	16	9,671	339	3,534	171	0.036	0.49	2.1e-05
V_EHF_01	25-ETS	10	6,130	205	1,938	91	0.027	0.51	1.1e-03
V_ELK1_01	25-ETS	16	11,809	351	4,178	151	0.027	0.29	8.1e-03

Nature Genetics: doi:10.1038/ng.3432

Motif	TF Cluster	Motif length (bp)	#SNPs	#Imbalanced SNPs	#SNPs in motif	#Imbalanced SNPs in Motif	Prop. substitutions in motif imbalanced	Enrich. imbalanced SNPs in motif	Q-value
V_ELK1_02	25-ETS	14	8,663	248	3,177	117	0.012	0.38	3.4e-03
V_ERG_01	25-ETS	9	4,040	160	1,335	91	0.042	0.83	3.0e-06
V_ETS2_B	25-ETS	14	13,382	453	4,019	171	0.042	0.34	2.4e-03
V_ETS_Q4	25-ETS	12	19,023	627	5,153	223	0.036	0.41	1.1e-04
V_ETS_Q6	25-ETS	8	15,045	450	2,884	115	0.031	0.43	5.0e-03
V_PU1_01	25-ETS	17	12,896	398	4,683	176	0.032	0.29	5.6e-03
ATF4_bZIP_1	49-CCAAT/CEBP	13	4,292	252	1,439	143	0.082	0.82	5.2e-09
Atf4.mouse_bZIP_1	49-CCAAT/CEBP	14	3,638	202	1,264	121	0.059	0.85	3.5e-08
CEBPB_bZIP_1	49-CCAAT/CEBP	10	1,757	165	600	124	0.15	1.3	5.6e-15
CEBPB_bZIP_2	49-CCAAT/CEBP	10	1,882	169	633	126	0.15	1.3	8.0e-16
Cebpb.mouse_bZIP_1	49-CCAAT/CEBP	10	1,828	174	625	133	0.14	1.4	8.1e-16
CEBPD_bZIP_1	49-CCAAT/CEBP	10	1,723	159	590	119	0.15	1.3	7.6e-14
CEBPE_bZIP_1	49-CCAAT/CEBP	10	1,938	176	640	128	0.16	1.3	4.4e-14
CEBPG_bZIP_1	49-CCAAT/CEBP	10	1,841	158	612	107	0.12	1.2	3.0e-10
CEBPG_bZIP_2	49-CCAAT/CEBP	10	1,968	171	643	123	0.16	1.3	2.1e-14
MA0102.1-Cebpa	49-CCAAT/CEBP	12	3,765	215	1,285	114	0.062	0.68	2.8e-06
NFIL3_bZIP_1	49-CCAAT/CEBP	12	1,971	141	650	85	0.077	0.97	1.1e-06
V_CEBPA_01	49-CCAAT/CEBP	14	4,164	201	1,446	106	0.041	0.64	2.3e-05
V_CEBPB_01	49-CCAAT/CEBP	14	7,013	320	2,297	158	0.056	0.63	1.5e-06
V_CEBPB_02	49-CCAAT/CEBP	14	3,951	227	1,367	142	0.067	0.93	2.0e-10
V_CEBP_C	49-CCAAT/CEBP	18	5,415	197	2,070	98	0.026	0.4	6.8e-03
V_CEBPDELTA_Q6	49-CCAAT/CEBP	12	3,881	190	1,290	107	0.067	0.81	3.5e-07
V_CEBP_Q2	49-CCAAT/CEBP	14	2,700	169	997	107	0.09	0.85	3.4e-07
V_CEBP_Q2_01	49-CCAAT/CEBP	12	5,281	250	1,566	127	0.061	0.83	1.1e-07
V_CEBP_Q3	49-CCAAT/CEBP	12	6,431	251	1,903	115	0.059	0.66	6.9e-06
V_VBP_01	49-CCAAT/CEBP	10	1,709	103	586	61	0.097	0.86	8.9e-05
Esrra_secondary	62.ESRRA	17	13,516	312	5,125	148	0.027	0.33	2.9e-03
Hnf4a_primary	62.ESRRA	17	10,853	238	4,186	127	0.021	0.48	2.6e-04
Rxra_primary	62.ESRRA	17	9,765	243	3,679	115	0.023	0.34	9.0e-03
V_LXR_DR4_Q3	62.ESRRA	16	7,524	168	2,877	89	0.023	0.48	2.4e-03
V_VDR_Q6	62.ESRRA	12	9,523	197	3,201	88	0.022	0.42	7.6e-03
NR2F1_nuclearreceptor_3	64.ESRRA/2	8	8,822	216	1,923	77	0.04	0.73	2.7e-04
NEUROG2_bHLH_2	65-NEURO/OLIG	10	4,621	187	1,227	73	0.071	0.58	2.9e-03
JDP2_bZIP_1	66-AP-1	9	5,028	401	969	210	0.27	1.7	3.4e-26
JDP2_bZIP_3	66-AP-1	9	6,834	561	1,207	281	0.3	1.8	8.3e-36
Jdp2.mouse_bZIP_1	66-AP-1	9	6,403	540	1,154	271	0.3	1.7	1.0e-31
Jundm2_secondary	66-AP-1	16	8,760	593	2,677	322	0.072	0.91	2.4e-23
MA0099.1-Fos	66-AP-1	8	7,725	569	1,257	273	0.28	1.8	7.1e-35
MA0099.2-AP1	66-AP-1	7	8,380	629	948	282	0.3	2.4	1.2e-43
MA0150.1-NFE2L2	66-AP-1	11	8,052	335	2,261	144	0.051	0.65	6.0e-06
NFE2_bZIP_1	66-AP-1	11	7,031	473	1,664	251	0.12	1.3	1.4e-22
V_AP1_01	66-AP-1	13	12,725	721	3,575	376	0.091	0.97	4.0e-25
V_AP1_C	66-AP-1	9	8,372	636	1,383	313	0.23	1.8	2.7e-39
V_AP1FJ_Q2	66-AP-1	11	8,468	378	2,321	180	0.067	0.85	6.2e-11
V_AP1_Q2	66-AP-1	11	8,245	427	2,192	208	0.074	0.94	9.1e-15
V_AP1_Q2_01	66-AP-1	12	9,975	527	2,818	262	0.059	0.88	3.0e-16
V_AP1_Q4	66-AP-1	11	9,258	473	2,352	231	0.077	1	3.0e-16
V_AP1_Q4_01	66-AP-1	8	9,698	632	1,504	294	0.21	1.8	1.3e-35
V_AP1_Q6	66-AP-1	11	11,269	661	2,659	337	0.1	1.2	2.8e-30
V_AP1_Q6_01	66-AP-1	9	10,862	698	2,065	346	0.19	1.6	6.8e-38
V_BACH1_01	66-AP-1	15	11,291	739	3,343	392	0.076	0.92	1.7e-27
V_BACH2_01	66-AP-1	11	9,879	638	2,635	338	0.1	1.1	5.9e-27
V_FRA1_Q5	66-AP-1	8	11,047	753	1,805	344	0.22	1.7	3.7e-39
V_MAF_Q6_01	66-AP-1	11	14,650	605	3,934	263	0.056	0.73	2.8e-13
V_NFE2_01	66-AP-1	11	10,290	543	2,674	261	0.086	0.96	7.5e-18
V_NRF2_Q4	66-AP-1	13	10,354	473	3,105	221	0.044	0.68	9.5e-10
V_TCF11MAFG_01	66-AP-1	22	7,415	342	3,100	173	0.034	0.29	5.9e-03
Atf1_primary	68-CREB/ATF	16	2,833	162	1,180	117	0.065	0.86	5.2e-09
ATF7_bZIP_1	68-CREB/ATF	14	2,551	158	999	113	0.1	0.95	5.3e-09
BATF3_bZIP_1	68-CREB/ATF	14	3,707	203	1,411	139	0.067	0.92	4.2e-10
CREB3_bZIP_1	68-CREB/ATF	14	4,232	199	1,685	142	0.056	0.9	1.8e-11
CREB3_bZIP_2	68-CREB/ATF	14	3,733	152	1,776	100	0	0.49	2.7e-04
Creb3l2.mouse_bZIP_1	68-CREB/ATF	13	6,186	204	2,520	106	0.006	0.36	5.3e-03
Creb5.mouse_bZIP_1	68-CREB/ATF	12	2,198	142	737	98	0.15	1.2	1.8e-10
JDP2_bZIP_2	68-CREB/ATF	12	2,907	177	956	118	0.14	1.1	5.3e-10
JDP2_bZIP_4	68-CREB/ATF	12	2,816	169	923	110	0.11	1.1	3.7e-10
Jdp2.mouse_bZIP_2	68-CREB/ATF	12	3,094	184	1,016	119	0.11	1.1	5.8e-10
Jundm2_primary	68-CREB/ATF	16	3,172	177	1,274	121	0.062	0.83	1.6e-08
MA0018.2-CREB1	68-CREB/ATF	8	3,047	145	894	90	0.091	1.2	1.1e-08
V_ATF_01	68-CREB/ATF	14	3,527	173	1,718	124	0.029	0.59	2.0e-06

Nature Genetics: doi:10.1038/ng.3432

Motif	TF Cluster	Motif length (bp)	#SNPs	#Imbalanced SNPs	#SNPs in motif	#Imbalanced SNPs in Motif	Prop. substitutions in motif imbalanced	Enrich. imbalanced SNPs in motif	Q-value
V_ATF3_Q6	68-CREB/ATF	14	8,251	256	3,107	125	0.032	0.39	1.9e-03
V_ATF4_Q2	68-CREB/ATF	12	6,717	236	2,181	119	0.04	0.66	2.8e-06
V_CREB_02	68-CREB/ATF	12	3,790	169	1,593	132	0.025	0.95	9.1e-12
V_CREB_Q2_01	68-CREB/ATF	14	4,918	215	2,109	154	0.041	0.79	3.0e-11
V_CREB_Q4_01	68-CREB/ATF	11	3,712	201	1,513	143	0.05	0.87	1.6e-10
V_CREM_Q6	68-CREB/ATF	11	6,658	262	2,480	159	0.034	0.74	5.6e-09
V_E4F1_Q6	68-CREB/ATF	10	2,763	131	1,167	99	0.042	0.9	2.3e-08
XBP1_bZIP_2	68-CREB/ATF	14	3,347	172	1,640	115	0.024	0.48	2.0e-04
V_COUP_DR1_Q6	85-NR/2	13	12,935	322	3,885	123	0.029	0.36	7.2e-03
SREBF2_bHLH_1	90-MYC/MAX	10	3,046	109	1,060	60	0.045	0.69	7.1e-04
TFEC_bHLH_1	90-MYC/MAX	10	5,125	167	1,659	77	0.029	0.53	3.3e-03
V_EBOX_Q6_01	90-MYC/MAX	10	12,604	336	3,754	135	0.034	0.45	7.7e-04
V_MAX_Q6	90-MYC/MAX	12	12,334	294	4,166	132	0.019	0.42	1.6e-03
V_SREBP1_01	90-MYC/MAX	11	5,369	175	1,746	78	0.034	0.47	5.6e-03
V_USF_Q6	90-MYC/MAX	10	9,613	257	3,257	112	0.031	0.37	6.9e-03
V_USF_Q6_01	90-MYC/MAX	12	13,214	322	4,585	142	0.018	0.36	3.6e-03
TBX15_TBX_2	91-T-box	8	6,960	167	1,688	64	0.039	0.68	1.7e-03
TBX5_TBX_1	91-T-box	8	6,688	173	1,615	70	0.044	0.77	3.2e-04
FIGLA_bHLH_1	109-Ebox/CACCTG	10	12,138	356	3,103	158	0.04	0.83	6.5e-09
ID4_bHLH_1	109-Ebox/CACCTG	10	15,409	401	3,753	159	0.032	0.73	1.8e-07
MESP1_bHLH_1	109-Ebox/CACCTG	10	15,259	373	4,086	160	0.028	0.7	1.7e-07
SNAI2_C2H2_1	109-Ebox/CACCTG	9	10,766	277	2,528	116	0.03	0.86	1.4e-07
TCF3_bHLH_1	109-Ebox/CACCTG	10	16,923	459	3,943	181	0.039	0.79	8.9e-10
TCF4_bHLH_1	109-Ebox/CACCTG	10	17,248	476	4,072	172	0.035	0.64	6.8e-07
TCF4_bHLH_2	109-Ebox/CACCTG	10	13,935	360	3,620	155	0.032	0.75	9.3e-08
Tcfe2a_primary	109-Ebox/CACCTG	17	17,451	437	6,199	207	0.019	0.43	3.7e-05
V_AREB6_01	109-Ebox/CACCTG	13	6,661	157	2,037	85	0	0.85	9.5e-06
V_AREB6_03	109-Ebox/CACCTG	12	12,165	319	3,752	135	0.034	0.47	5.2e-04
V_E47_02	109-Ebox/CACCTG	16	15,052	401	4,912	180	0.026	0.47	2.0e-05
V_AREB6_02	110-Ebox/CACCTn	12	7,648	199	2,206	96	0.016	0.77	2.1e-05
V_DELTAEF1_01	110-Ebox/CACCTn	11	7,981	198	2,179	98	0.03	0.89	1.4e-06
Ascl2.mouse_bHLH_1	112-Ebox	10	15,961	481	3,570	143	0.026	0.43	2.3e-03
MA0048.1-NHLH1	112-Ebox	12	17,710	472	5,033	168	0.022	0.33	3.5e-03
MA0091.1-TAL1::TCF3	112-Ebox	12	8,013	316	2,276	117	0.04	0.4	5.2e-03
MSC_bHLH_1	112-Ebox	10	10,492	314	2,447	106	0.036	0.55	6.7e-04
Myf6_primary	112-Ebox	16	16,776	423	5,573	172	0.025	0.3	5.7e-03
NHLH1_bHLH_2	112-Ebox	10	14,459	398	3,382	128	0.028	0.47	7.7e-04
Tcf21.mouse_bHLH_1	112-Ebox	14	10,151	334	3,057	135	0.033	0.44	9.3e-04
TFAP4_bHLH_1	112-Ebox	10	8,829	298	2,019	97	0.04	0.53	9.3e-04
TFAP4_bHLH_2	112-Ebox	10	9,361	311	2,104	96	0.036	0.48	4.8e-03
V_AP4_Q6_01	112-Ebox	9	19,860	551	4,432	166	0.032	0.45	4.4e-04
V_E12_Q6	112-Ebox	11	18,644	511	4,884	195	0.033	0.56	1.2e-06
V_E2A_Q2	112-Ebox	14	24,219	608	7,812	233	0.024	0.26	5.5e-03
V_E2A_Q6	112-Ebox	8	15,113	428	3,231	155	0.033	0.79	3.5e-08
V_E47_01	112-Ebox	15	27,640	691	9,173	291	0.028	0.35	2.0e-05
V_HEN1_02	112-Ebox	22	23,761	572	9,525	266	0.021	0.22	6.0e-03
V_LBP1_Q6	112-Ebox	7	13,067	406	2,277	101	0.038	0.53	1.4e-03
V_LMO2COM_01	112-Ebox	12	21,490	552	6,080	205	0.028	0.4	2.0e-04
V_MATH1_Q2	112-Ebox	10	17,945	515	4,123	169	0.038	0.53	6.6e-05
V_MYOD_01	112-Ebox	12	14,414	448	4,260	173	0.034	0.4	4.0e-04
V_MYOD_Q6	112-Ebox	10	18,128	518	4,080	191	0.036	0.74	1.8e-08
V_MYOD_Q6_01	112-Ebox	18	33,359	777	12,497	357	0.023	0.3	2.8e-05
V_MYOGENIN_Q6	112-Ebox	8	17,084	493	3,376	159	0.042	0.73	2.7e-07
V_TAL1ALPHAE47_01	112-Ebox	16	15,527	499	5,239	210	0.035	0.33	8.5e-04
V_TAL1BETAE47_01	112-Ebox	16	13,086	414	4,526	174	0.03	0.29	5.3e-03
V_TAL1BETAITF2_01	112-Ebox	16	12,090	408	3,988	181	0.034	0.44	4.0e-05
V_TAL1_Q6	112-Ebox	10	19,442	575	4,574	174	0.033	0.38	2.3e-03
ZNF238_C2H2_1	112-Ebox	13	11,645	363	3,405	134	0.033	0.35	5.7e-03
GLI2_C2H2_1	116-GLI	14	18,790	444	6,813	193	0.024	0.27	6.6e-03
V_GLI1_Q2	116-GLI	10	16,534	388	4,548	136	0.025	0.36	6.8e-03
V_GLI_Q2	116-GLI	12	16,562	385	5,347	154	0.024	0.32	6.5e-03
ZBTB7A_C2H2_1	116-GLI	12	10,208	250	3,699	128	0.028	0.51	1.3e-04
ZBTB7B_C2H2_1	116-GLI	12	10,493	231	3,834	124	0.026	0.57	2.6e-05
Zbtb7b_secondary	116-GLI	17	9,597	197	3,745	110	0.019	0.53	1.9e-04
ZBTB7C_C2H2_1	116-GLI	12	12,665	297	4,616	149	0.021	0.47	1.3e-04
GLIS2_C2H2_1	117-GLI/2	14	19,896	438	8,111	223	0.021	0.33	2.6e-04
Glis2_primary	117-GLI/2	16	14,536	331	6,315	187	0.025	0.39	7.7e-05
GLIS3_C2H2_1	117-GLI/2	14	13,128	277	5,240	139	0.022	0.34	5.3e-03
V_LRF_Q2	117-GLI/2	9	15,880	342	4,354	128	0.027	0.46	8.1e-04
V_ZNF515_01	117-GLI/2	10	20,598	469	6,709	192	0.024	0.34	1.2e-03

Nature Genetics: doi:10.1038/ng.3432

Motif	TF Cluster	Motif length (bp)	#SNPs	#Imbalanced SNPs	#SNPs in motif	#Imbalanced SNPs in Motif	Prop. substitutions in motif imbalanced	Enrich. imbalanced SNPs in motif	Q-value
Zfp281_secondary	117-GLI/2	17	9,254	223	3,847	121	0.022	0.4	1.5e-03
ZIC1_C2H2_1	117-GLI/2	14	22,509	486	8,260	218	0.022	0.3	1.2e-03
ZIC3_C2H2_1	117-GLI/2	15	23,025	496	8,767	229	0.024	0.28	1.9e-03
Zic3.mouse_C2H2_1	117-GLI/2	15	24,615	549	9,349	251	0.025	0.27	1.6e-03
ZIC4_C2H2_1	117-GLI/2	15	25,890	558	9,790	254	0.024	0.27	1.6e-03
Sfpi1_secondary	125.SFPI1	14	3,348	110	1,242	63	0.028	0.65	1.3e-03
MA0061.1-NF-kappaB	134-NF-kB	10	9,527	217	2,310	79	0.027	0.6	1.3e-03
V_CREL_01	134-NF-kB	10	7,987	221	2,070	80	0.026	0.5	4.8e-03
V_NFKAPPAB_01	134-NF-kB	10	9,830	223	2,386	79	0.025	0.56	3.7e-03
V_NFKAPPAB65_01	134-NF-kB	10	7,521	217	1,806	82	0.042	0.68	4.9e-04
Klf12.mouse_C2H2_1	135-KLF/Sp4	15	19,055	474	6,677	200	0.02	0.28	4.7e-03
KLF14_C2H2_1	135-KLF/Sp4	14	20,632	440	7,424	219	0.018	0.48	8.5e-07
SP4_C2H2_1	135-KLF/Sp4	17	25,274	594	9,756	285	0.019	0.32	1.5e-04
Sp4_secondary	135-KLF/Sp4	15	13,661	308	4,932	140	0.02	0.34	6.1e-03
EGR1_C2H2_1	136-EGR/KLF/Sp1	14	19,050	399	7,450	200	0.019	0.37	1.8e-04
Egr1_primary	136-EGR/KLF/Sp1	14	28,075	576	10,764	284	0.021	0.37	6.9e-06
EGR2_C2H2_2	136-EGR/KLF/Sp1	15	19,136	409	7,540	201	0.017	0.33	7.7e-04
EGR3_C2H2_1	136-EGR/KLF/Sp1	15	21,796	457	8,651	227	0.022	0.33	3.0e-04
KLF16_C2H2_1	136-EGR/KLF/Sp1	11	33,160	742	9,849	309	0.02	0.5	1.4e-08
Klf7_primary	136-EGR/KLF/Sp1	16	33,502	759	12,209	338	0.021	0.3	8.1e-05
MA0039.2-Klf4	136-EGR/KLF/Sp1	10	33,053	750	8,600	268	0.029	0.47	4.3e-07
MA0162.1-Egr1	136-EGR/KLF/Sp1	11	14,146	312	4,916	150	0.024	0.48	9.4e-05
SP1_C2H2_1	136-EGR/KLF/Sp1	11	28,598	709	8,150	288	0.029	0.53	1.6e-08
SP3_C2H2_1	136-EGR/KLF/Sp1	11	33,581	737	10,180	305	0.024	0.46	2.3e-07
Sp4_primary	136-EGR/KLF/Sp1	17	36,321	736	14,116	366	0.018	0.36	2.0e-07
SP8_C2H2_1	136-EGR/KLF/Sp1	12	29,392	670	9,716	300	0.021	0.45	2.3e-07
V_EGR1_01	136-EGR/KLF/Sp1	12	12,686	276	4,472	126	0.027	0.38	2.6e-03
V_EGR_Q6	136-EGR/KLF/Sp1	11	28,407	577	8,933	229	0.02	0.34	3.0e-04
V_GKLF_02	136-EGR/KLF/Sp1	12	34,418	758	10,509	312	0.024	0.44	1.3e-07
V_SP1_Q2_01	136-EGR/KLF/Sp1	10	45,118	950	12,276	352	0.024	0.46	4.3e-08
V_SP1_Q4_01	136-EGR/KLF/Sp1	13	50,952	1,022	16,375	432	0.022	0.4	4.2e-09
V_SP1_Q6	136-EGR/KLF/Sp1	13	46,390	902	15,140	381	0.02	0.38	9.2e-08
V_SP1_Q6_01	136-EGR/KLF/Sp1	10	41,300	844	10,906	307	0.023	0.47	9.2e-08
MA0079.2-SP1	137-Sp1	10	46,806	900	12,661	315	0.021	0.38	1.2e-05
V_CKROX_Q2	137-Sp1	9	43,237	852	10,563	247	0.023	0.25	5.7e-03
V_SP1_01	137-Sp1	10	29,246	631	8,206	221	0.023	0.33	1.1e-03
V_SP1_02	137-Sp1	11	43,762	818	12,538	283	0.018	0.28	1.3e-03
V_SP2_01	137-Sp1	9	16,849	318	4,580	136	0.023	0.67	3.9e-06
V_SP4_Q5	137-Sp1	11	45,442	883	13,434	338	0.019	0.38	3.1e-06
V_ZBP89_Q4	137-Sp1	10	33,387	750	9,575	274	0.025	0.36	8.4e-05
CTCF_Upstream	138-CTCF_Upstream	35	57,404	1,282	29,952	780	0.022	0.23	6.6e-08
MA0050.1-IRF1	145-IRF	12	6,418	235	1,751	87	0.032	0.46	5.1e-03
V_ICSBP_Q6	145-IRF	12	6,130	211	1,740	85	0.04	0.53	1.8e-03
V_KROX_Q6	151-ZNF	14	41,735	848	14,742	355	0.02	0.25	6.6e-04
V_MAZR_01	151-ZNF	13	37,555	756	12,619	298	0.023	0.24	3.5e-03
V_MOVOB_01	151-ZNF	7	17,939	376	5,521	154	0.027	0.42	6.6e-04
V_SP1SP3_Q4	151-ZNF	11	31,777	628	9,856	255	0.024	0.4	4.1e-05
V_WT1_Q6	151-ZNF	9	33,347	637	9,706	244	0.022	0.41	4.2e-05
V_ZNF219_01	151-ZNF	12	38,476	749	12,901	322	0.023	0.37	2.0e-06
Zfp281_primary	151-ZNF	15	39,835	837	14,274	361	0.022	0.27	1.9e-04
Zfp740.mouse_C2H2_1	151-ZNF	10	19,902	469	6,944	199	0.026	0.29	3.7e-03
Zfp740_primary	151-ZNF	16	27,755	591	11,243	289	0.021	0.28	2.7e-04
ZNF740_C2H2_1	151-ZNF	10	22,545	492	7,591	217	0.025	0.4	4.7e-05
ZNF740_C2H2_2	151-ZNF	10	27,879	625	8,752	266	0.027	0.45	1.9e-06
IRF3_IRF_1	157-IRF/2	21	9,076	268	3,647	135	0.026	0.34	5.5e-03
IRF7_IRF_1	157-IRF/2	14	4,436	149	1,420	78	0.05	0.74	1.1e-04
IRF8_IRF_1	157-IRF/2	14	4,477	164	1,494	79	0.049	0.55	1.3e-03
IRF8_IRF_2	157-IRF/2	14	3,996	151	1,323	81	0.056	0.73	5.0e-05
IRF9_IRF_1	157-IRF/2	15	4,732	177	1,607	87	0.053	0.56	6.9e-04
Irf3_primary	158-IRF/3	14	5,622	160	2,003	79	0.032	0.49	3.5e-03
Isgf3g_primary	158-IRF/3	15	3,072	116	1,074	64	0.042	0.69	6.0e-04
Zic1_primary	177-Zic	14	21,194	451	7,880	218	0.025	0.39	4.3e-05
Zic2_primary	177-Zic	15	20,559	426	7,865	212	0.024	0.39	3.5e-05
Zic3_primary	177-Zic	15	21,713	455	8,336	223	0.024	0.36	6.6e-05
MA0116.1-Zfp423	178.ZFP423	15	12,533	302	4,658	142	0.028	0.35	3.2e-03
TEAD1_TEA_1	180-TEAD/TEF	10	6,989	351	1,762	141	0.069	0.72	2.0e-07
TEAD3_TEA_1	180-TEAD/TEF	17	5,325	236	1,911	114	0.05	0.45	1.6e-03
TEAD3_TEA_2	180-TEAD/TEF	8	6,262	327	1,285	134	0.095	1.1	7.8e-11
TEAD4_TEA_1	180-TEAD/TEF	10	6,999	350	1,654	132	0.059	0.72	3.6e-06
V_TEF1_Q6_03	180-TEAD/TEF	9	8,031	330	1,971	119	0.04	0.58	1.6e-04

Nature Genetics: doi:10.1038/ng.3432

Motif	TF Cluster	Motif length (bp)	#SNPs	#Imbalanced SNPs	#SNPs in motif	#Imbalanced SNPs in Motif	Prop. substitutions in motif imbalanced	Enrich. imbalanced SNPs in motif	Q-value
MA0090.1-TEAD1	181-TEAD/TEF/2	12	8,778	356	2,633	143	0.042	0.44	5.9e-04
V_TEF_01	181-TEAD/TEF/2	12	8,778	356	2,633	143	0.042	0.44	7.2e-04
V_SREBP2_Q6	187.PAX4/4	12	22,340	478	7,567	202	0.025	0.33	9.7e-04
MA0002.1-RUNX1	188-RUNX/AML	11	4,198	237	1,180	101	0.057	0.65	1.0e-04
MA0002.2-RUNX1	188-RUNX/AML	11	14,386	556	3,817	180	0.038	0.3	6.8e-03
RUNX2_RUNX_3	188-RUNX/AML	9	3,710	207	1,079	83	0.077	0.5	4.8e-03
RUNX3_RUNX_2	188-RUNX/AML	10	4,688	245	1,386	100	0.067	0.5	1.4e-03
RUNX3_RUNX_4	188-RUNX/AML	10	5,402	281	1,535	109	0.056	0.48	1.7e-03
V_AML1_Q4	188-RUNX/AML	7	5,360	296	1,055	89	0.11	0.66	5.4e-04
V_AML2_01	188-RUNX/AML	8	5,310	275	1,347	98	0.08	0.52	1.3e-03
V_AML_Q6	188-RUNX/AML	15	12,305	504	4,179	203	0.038	0.26	6.9e-03
Bcl6b_secondary	193.ASCL2	16	37,847	702	14,243	309	0.019	0.23	2.1e-03
Smad3_secondary	193.ASCL2	17	29,458	531	12,094	264	0.015	0.28	5.9e-04
Tcfap2a.mouse_TFAP_3	202-AP-2	13	18,510	391	6,210	161	0.024	0.3	5.0e-03
TFAP2A_AP2_6	202-AP-2	13	17,998	378	6,076	156	0.024	0.3	8.4e-03
TFAP2A_TFAP_3	202-AP-2	13	20,061	420	6,726	173	0.024	0.3	5.3e-03
TFAP2B_TFAP_3	202-AP-2	13	17,250	348	5,789	144	0.022	0.31	8.5e-03
Tcfap2c_secondary	203-AP-2/2	14	15,208	296	5,281	129	0.017	0.34	6.5e-03
V_AP2_Q6	203-AP-2/2	12	30,189	531	10,524	224	0.018	0.28	2.7e-03
V_AP2_Q6_01	203-AP-2/2	13	45,239	782	15,090	303	0.02	0.22	3.9e-03
Rfx2.mouse_RFX_1	205-RFX	16	3,277	129	1,191	69	0.028	0.59	1.4e-03
RFX2_RFX_1	205-RFX	16	2,957	126	1,038	65	0.05	0.59	3.3e-03
Rfx3.mouse_RFX_1	205-RFX	16	3,470	131	1,272	73	0.038	0.63	6.8e-04
RFX4_RFX_1	205-RFX	16	3,808	144	1,409	73	0.041	0.48	5.0e-03
RFX5_RFX_2	205-RFX	16	4,270	162	1,593	82	0.041	0.46	3.4e-03
RFX5_RFX_3	205-RFX	16	4,270	162	1,593	82	0.041	0.46	3.1e-03
V_TAXCREB_02	211-CREB/ATF/2	15	7,691	297	2,910	150	0.04	0.44	2.6e-04
E2F2_primary	220-E2F/3	15	6,815	107	2,478	56	0	0.54	8.8e-03
E2F3_primary	220-E2F/3	15	6,173	108	2,258	62	0.016	0.66	1.5e-03
Plagl1_primary	221.GCM1/3	16	20,448	369	7,532	170	0.012	0.33	2.4e-03
MA0119.1-TLX1::NFIC	222-CTF/NF-I	14	14,006	561	4,451	311	0.042	0.85	8.1e-19
NFIA_NFI_1	222-CTF/NF-I	15	14,083	678	4,624	419	0.071	0.98	8.5e-31
NFIB_NFI_1	222-CTF/NF-I	15	14,197	678	4,657	419	0.063	0.98	7.9e-29
NFIX_NFI_1	222-CTF/NF-I	15	16,685	743	5,434	452	0.051	0.96	7.7e-34
NFIX_NFI_4	222-CTF/NF-I	15	13,868	625	4,486	386	0.042	1	1.0e-29
V_CTF1_01	222-CTF/NF-I	14	14,006	561	4,451	311	0.042	0.85	1.2e-18
V_NF1_Q6	222-CTF/NF-I	18	22,387	753	8,433	402	0.026	0.52	6.0e-13
V_NF1_Q6_01	222-CTF/NF-I	17	17,226	624	6,251	353	0.036	0.67	5.2e-16
Hic1.mouse_C2H2_1	239.HIC1/2	18	19,648	447	7,799	217	0.023	0.3	1.2e-03
CTCF_C2H2_1	251-CTCF_Core/2	17	12,530	455	4,431	286	0.059	0.87	3.1e-19
CTCF_Core	251-CTCF_Core/2	35	72,285	1,543	36,762	944	0.023	0.27	6.3e-13
MA0139.1-CTCF	251-CTCF_Core/2	19	30,875	873	11,777	498	0.042	0.6	3.1e-19
V_CTCF_02	251-CTCF_Core/2	20	35,069	955	13,799	543	0.036	0.55	4.5e-18

Nature Genetics: doi:10.1038/ng.3432

Supplementary Table 17. Features for prediction of regulatory variation impacting TF occupancy

Category	Feature	Average coefficient
	Intercept	-4.865
Covariates	Read depth	1.020
	Num. hets^1	-2.281
	Num. hets^2	1.097
Site-dependent	MCV^1	0.316
	MCV^2	-0.375
	CpG Island *	-0.479
	3' UTR *	-0.039
	coding *	-0.016
	intron *	0.151
	intergenic *	0.148
	Dist. to TSS^1	0.010
	Dist. to TSS^2	-0.041
	DHS strength^1	-0.018
	DHS strength^2	0.043
	Width of DHS	-0.360
	#nearby binding sites^1	-0.442
	#nearby binding sites^2	0.273
	PhastCons *	-0.047
	Footprint presence *	0.735
	Footprint occupancy score	-0.243
TF-dependent	log(FIMO score)^1	0.481
	log(FIMO score)^2	-0.343
	logodds difference	0.094
	position (mean) *	0.620

List of features used in logistic regression model, divided into covariates used for training, site-dependent features specific to each SNV, and features specific to each overlapping TF recognition sequence (**Fig. 7a**, **Online Methods**). Asterisk indicates binary factors. Shown are medians of standardized coefficients across all motifs. See also **Supplementary Fig. 13a**.

Nature Genetics: doi:10.1038/ng.3432

SUPPLEMENTAL REFERENCES

64. Stergachis, A. B. *et al.* Developmental Fate and Cellular Maturity Encoded in Human Regulatory DNA Landscapes. *Cell* **154,** 888–903 (2013).
65. Vierstra, J., Wang, H., John, S., Sandstrom, R. & Stamatoyannopoulos, J. A. Coupling transcription factor occupancy to nucleosome architecture with DNase-FLASH. *Nat. Methods* **11,** 66–72 (2014).
66. Kellis, M. *et al.* Reply to Brunet and Doolittle: Both selected effect and causal role elements can influence human biology and disease. *Proc. Natl. Acad. Sci. U.S.A.* **111,** E3366 (2014).
67. Vierstra, J. *et al.* Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science* **346,** 1007–1012 (2014).
68. Maurano, M. T. *et al.* Role of DNA Methylation in Modulating Transcription Factor Occupancy. *Cell Rep* **12,** 1184–1195 (2015).

Nature Genetics: doi:10.1038/ng.3432