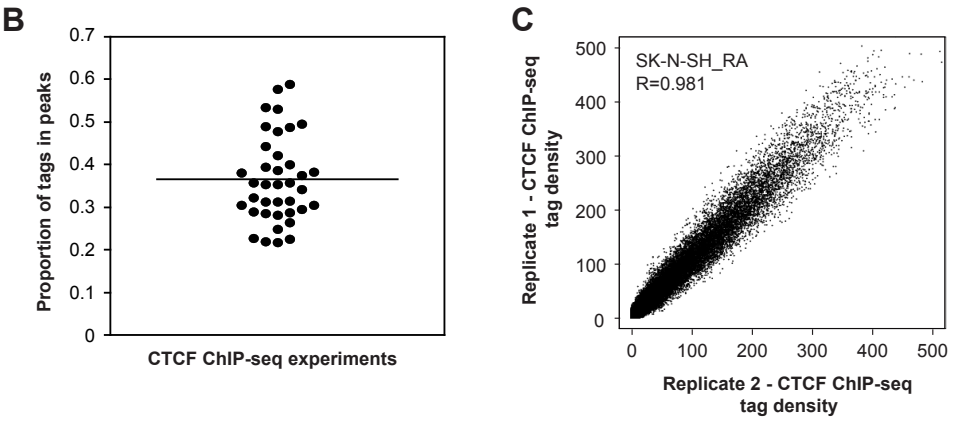
SUPPLEMENTARY FIGURES

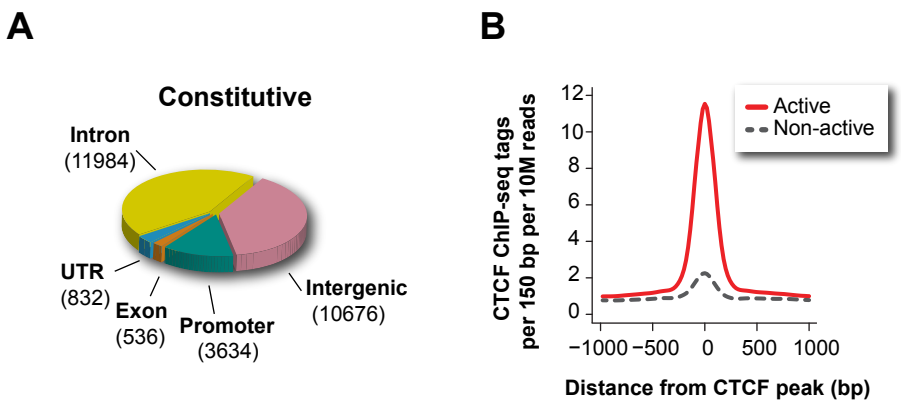
A

Cell type	Description	# of non-redundant tags	# of peaks	Pearson correlation (R) between replicates	Peak concodance between replicates
AG09309	toe skin fibroblast	21,910,706	50,269	0.911	0.903
AG09319	gum fibroblast	38,692,770	54,617	0.966	0.952
AG10803	abdomen fibroblast	48,484,390	53,560	0.960	0.947
AoAF	aortic adventitial fibroblast	33,811,909	51,475	0.912	0.945
BJ	foreskin fibroblast	20,905,180	53,215	0.971	0.938
Caco-2	colorectal adenocarcinoma	23,723,539	54,585	0.905	0.936
GM06990	lymphoblast	19,444,414	51,373	0.940	0.937
HBMEC	brain microvascular endothelium	30,872,109	62,983	0.904	0.918
HEEpic	esophageal epithelium	29,927,913	51,375	0.894	0.902
HeLa-S3	cervical carcinoma	13,947,609	53,551	0.924	0.870
HepG2	hepatocellular carcinoma	18,748,465	60,013	0.842	0.886
HMF	mammary fibroblast	28,508,296	58,975	0.950	0.931
HPAF	pulmonary artery fibroblast	45,238,597	59,574	0.979	0.965
HPF	pulmonary fibroblast	41,255,680	53,024	0.947	0.947
HRE	renal epithelium	13,299,282	55,647	0.936	0.924
K562	chronic myeloid leukemia	21,326,777	50,948	0.910	0.884
SAEC	small airway epithelium	14,539,607	54,104	0.953	0.925
SK-N-SH_RA	neuroblastoma	17,711,375	56,911	0.981	0.939
WERI-Rb-1	retinoblastoma	18,682,528	58,468	0.927	0.912

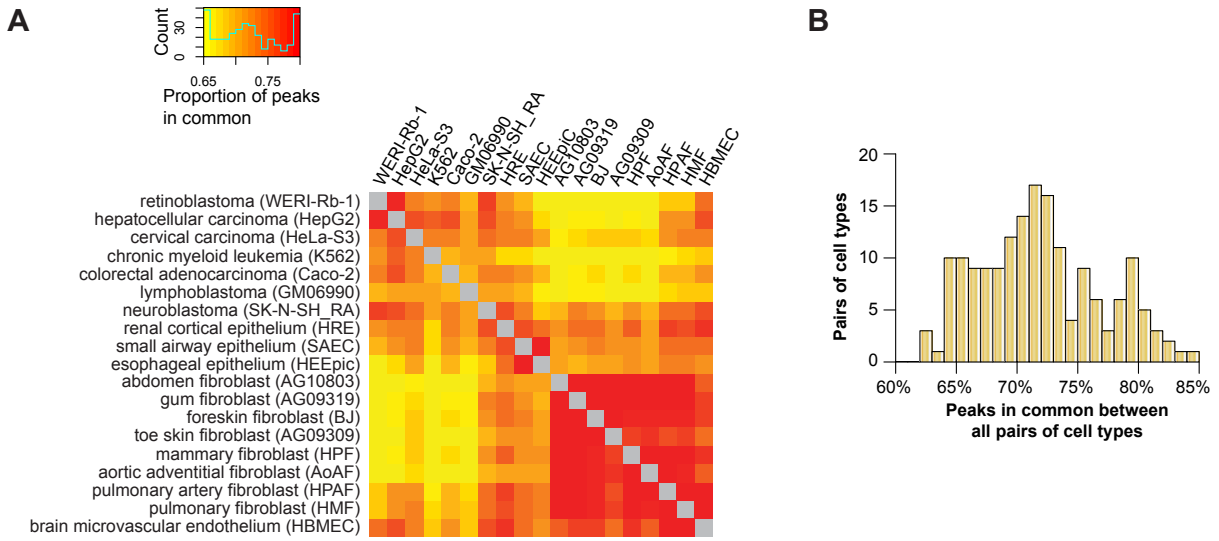
Total tags = 501,076,146



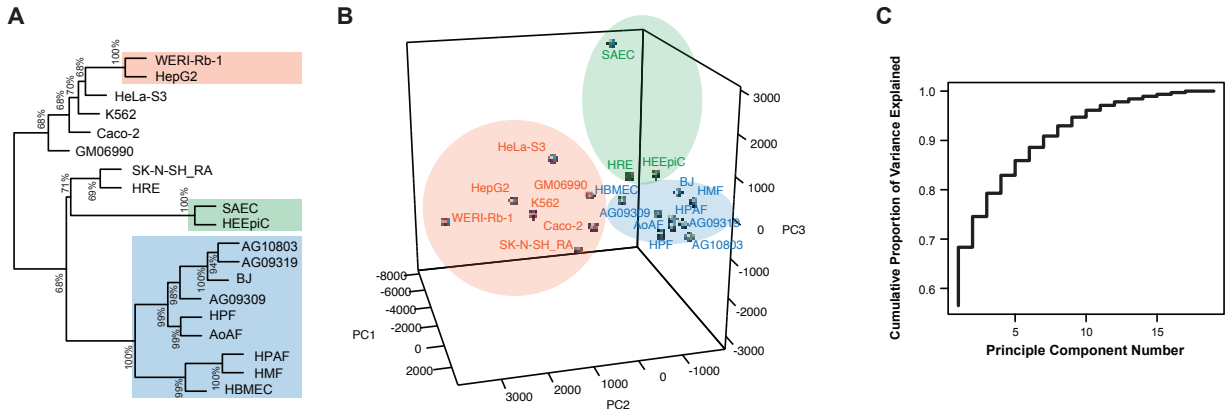
Supplementary Figure S1. Summary of CTCF ChIP-seq experiments for 19 surveyed cell types. (A) Number of tags represents the total uniquely-mapped reads (excluding PCR duplicates) for both biological replicates of each cell type. Number of peaks refers to the number of CTCF binding sites considered present in a given cell type (see Methods). Pearson correlation coefficient (R) measures concordance between replicate density tracks on chromosome 19 (average=0.93). Peak concordance between replicates refers to the average of the proportion of peaks in one replicate that are found in the second replicate and vice versa. (B) SPOT scores (Signal Portion of Tags, proportion of ChIP-seq tags located within hotspots) for 19 ChIP-seq experiments, showing degree of enrichment. Y-axis indicates proportion of all sequenced tags mapping to hotspot regions. (C) Example of Pearson correlation R for two K562 replicates. X-axis, Y-axis, replicate 1 and 2 tag densities on chr19.



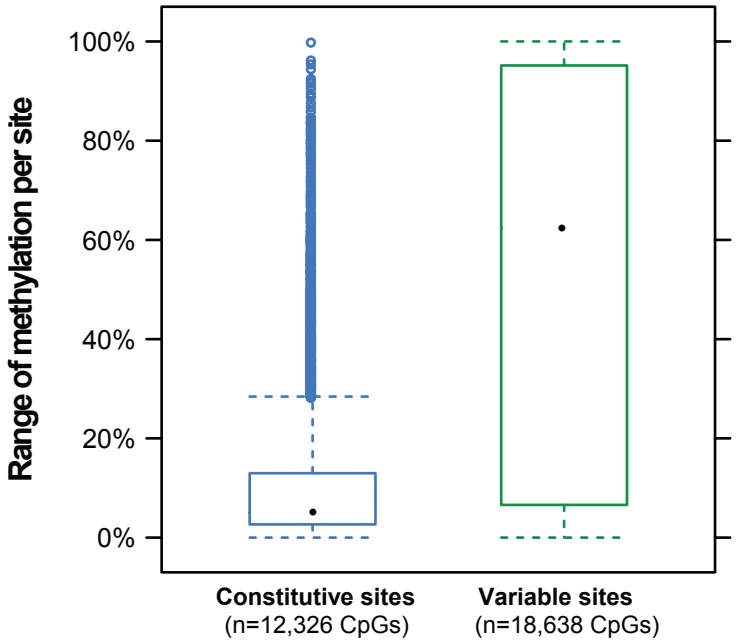
Supplementary Figure S2. CTCF *in vivo* occupancy exhibits marked plasticity. (A) Genomic distribution of constitutive sites (Gencode V7 annotations). Promoter, 1 kb region upstream of transcription start site. (B) Aggregate normalized CTCF signal in active (red) and non-active (grey) cell types. Vertical axis, normalized mean ChIP-seq tag density. Horizontal axis, distance from peak center.



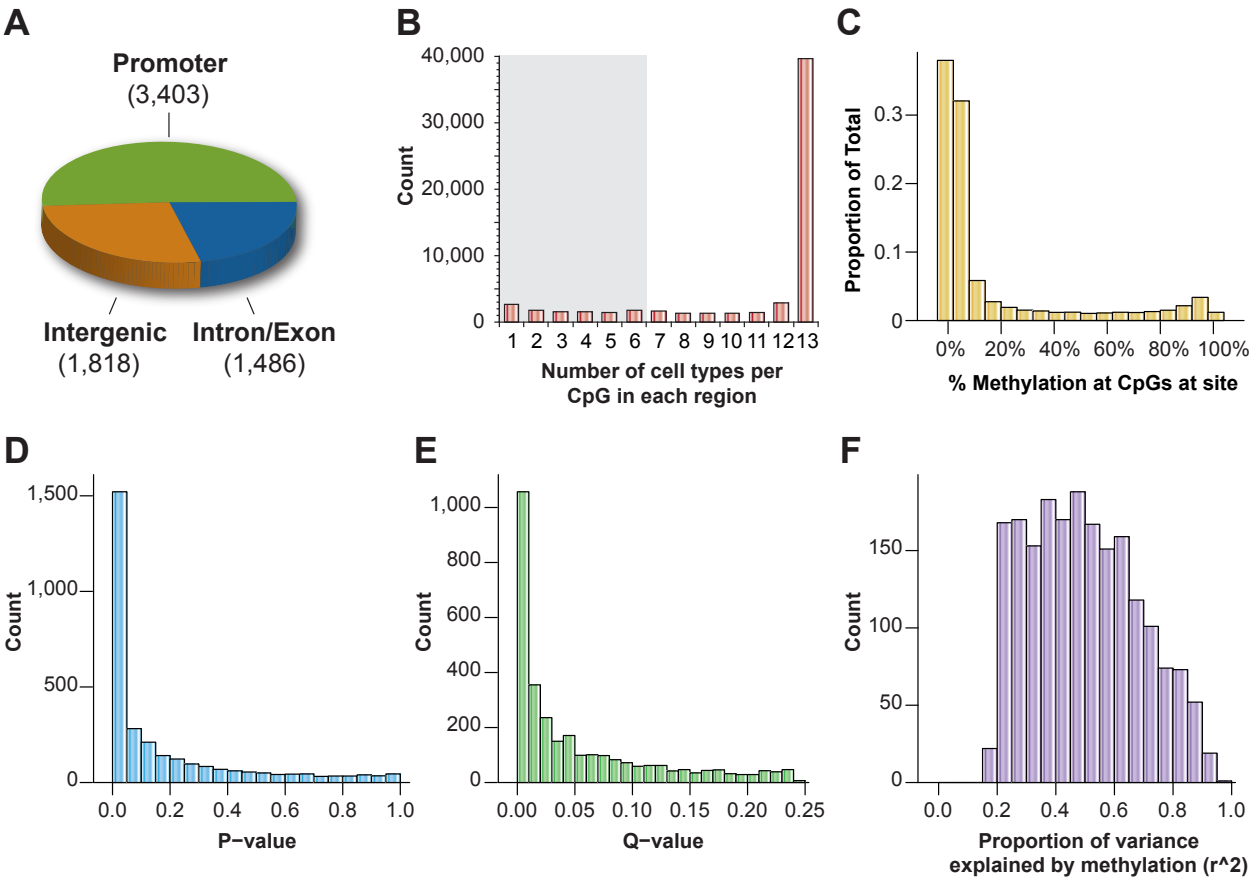
Supplementary Figure S3. Degree of overlapping binding landscape. (A) Overlap between pairs of cell types, in terms of proportion of peaks in common. (B) Percentage of binding sites shared between all possible pairs of individual replicate samples.



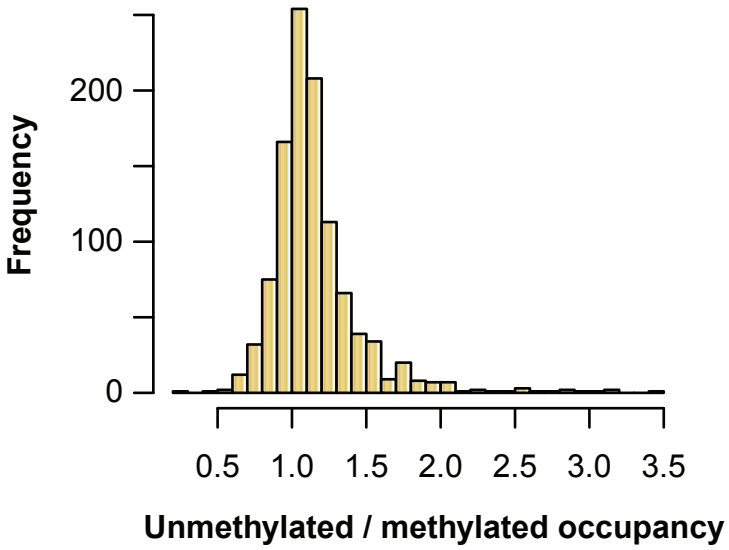
Supplementary Figure S4. CTCF occupancy landscape distinguishes similar cell types. (A) Bootstrap confidence values of hierarchical clustering analysis of CTCF binding sites. Percentages indicate approximately unbiased (AU) p-values; boxes indicate significant clusters. (B) Principle components analysis separation of cell types based on CTCF occupancy. (C) Cumulative proportion of variance explained by each principle component.



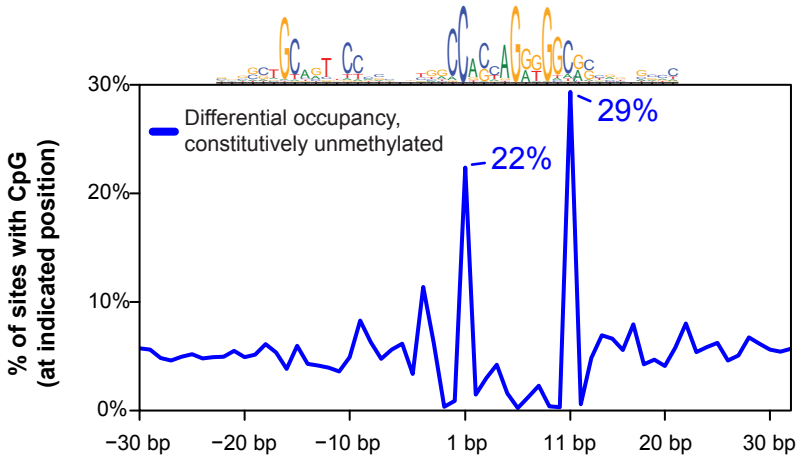
Supplementary Figure S5. Overall methylation variability across constitutive and variable CTCF sites in 13 cell types. Y-axis denotes the range between the most- and least-methylated cell types at each CpG overlapping constitutive and variable CTCF sites.



Supplementary Figure S6. Statistical association of variable methylation with differential occupancy. (A) Location relative to genes of sites surveyed by reduced representation bisulfite sequencing, which enriches for genic and CpG-island regions, including promoters. (B) Most CpGs were surveyed in all 13 samples. Sites without with at least one surveyed CpG for at least 6 samples (gray shading) were excluded from association analysis. (C) Methylation levels (X-axis) observed across all CTCF sites in all cell types; most CTCF sites are unmethylated. (D-E) Association of methylation with ChIP-seq occupancy identifies 905, 1,677 and 2,046 significant sites at FDR levels 1%, 5% and 10%, respectively. Histogram of p-values (D) and FDR-adjusted q-values (E) for all tested binding sites. (F) Effect size of methylation differences associated with differences in occupancy (FDR 5%), measured by r^2 of a linear regression



Supplementary Figure S7. Overall relationship of CTCF occupancy with methylation at 1,076 sites without a significant association. X-axis denotes ratio of average occupancy in unmethylated cell types divided by the average in methylated cell types.



Supplementary Figure S8. Variable CTCF sites without methylation differences. Frequency of a CpG (Y-axis) at positions relative to the CTCF motif (X-axis) shown for sites with differential binding but no differential methylation (blue). Note that the presence of a CpG at positions 1 and 11 is similar to that at sites where the variable methylation was associated with occupancy (Figure 4, red), but sites with variable methylation that is not associated with occupancy (Figure 4, gray) are depleted for these CpGs.

SUPPLEMENTARY TABLES

Table S1

Cell	Description	Medium	Supplements		
			%FBS	NEAA[A]	NaPyruvate[B]
AG09309	adult toe fibroblast	MEM[C]	15%	√	
AG09319	adult gum fibroblast	MEM[C]	15%	√	
AG10803	adult abdomen fibroblast	MEM[C]	15%	√	
AoAF	human aortic adventitial fibroblast	Lonza SCBM basal medium+SCGM SingleQuot Kit[D]			
BJ	skin fibroblast	MEM[C]	10%	√	√
Caco-2	colorectal adenocarcinoma	MEM[C]	20%	√	√
GM06990	lymphoblast	RPMI[E]	15%		
HBMEC	human brain microvascular endothelium	Endothelial cell medium (ECM)[F]			
HEEpiC	human esophageal epithelium	Epithelial cell medium-2 (EpiCM2)[F]			
HMF	human mammary fibroblast	Fibroblast medium (FM)[F]			
HPAF	human pulmonary artery fibroblast	Fibroblast medium (FM)[F]			
HPF	human pulmonary fibroblast	Fibroblast medium (FM)[F]			
HRE	human renal epithelium	REBM Basal Medium+REGM SingleQuot Kit[D]			
HeLa-S3	cervical adenocarcinoma	DMEM[G]	10%		
HepG2	hepatocellular carcinoma	MEM[C]	10%	√	√
K562	chronic myeloid leukemia	IMDM[H]	10%		
SAEC	human small airway epithelium	Lonza SABM Basal Medium+SAGM SingleQuot Kit[D]			
SK-N-SH_RA[I]	neuroblastoma	RPMI[E]	10%		√
WERI-Rb-1	retinoblastoma	RPMI[E]	10%		

Supplementary Table S1. Growth conditions for immortal cell lines and primary cells

All cells are supplemented with 25 IU/L penicillin and 25 mg/L streptomycin (Invitrogen)

Supplements: A: **NEAA**=0.1 mM non-essential amino acids (Invitrogen); B: **NaPyruvate**=1 mM sodium pyruvate (Invitrogen)

Media: C: **MEM**=Eagle's minimum essential medium with Earle's salts and 2 mM L-Glutamine (Cellgro); D: Lonza medium and supplements; E: **RPMI**= RPMI 1640 Medium with 2 mM L-Glutamine (Cellgro); F: ScienCell medium; G: **DMEM**=Dulbecco's Modification of Eagle's Medium with 4.5 g/L Glucose, 110 mg/L Sodium Pyruvate and 4mM L-Glutamine (Cellgro); H: **IMDM**=Iscove's modification of DMEM with L-glutamine and 25 mM HEPES, without α-thioglycerol and β-mercaptoethanol (Cellgro)

I: At 60% confluency, treat with 6 μM retinoic acid for 48 hours to differentiate

Supplementary Table S2. Location of ChIP-seq binding positions in 19 cell lines.
chrom, chromStart, chromEnd, ID, strand. MCV (for multiple cell verified): the number of cell types in which the site is occupied. "cell name_pk": "1" and "0" for presence and absence of the occupancy, respectively. "cell name_density": the normalized (based on 10 million mapped reads) maximum density value within the CTCF peak in a given cell line.

Supplementary Table S3. Sites characteristic of three groups from dendrogram.
chrom, chromStart, chromEnd hg19 coordinates of CTCF binding sites displaying differential CTCF occupancy and the binding pattern is denoted by 4th column as immortal-up-regulated, immortal-down-regulated, epithelium-up-regulated, epithelium-down-regulated, fibroblast-up-regulated, and fibroblast-down-regulated.

Supplementary Table S4. Methylation calls in CTCF sites used for 13 cell types.
chrom, chromStart, chromEnd ID, Strand, the 0-indexed hg19 location of the CpG; coordinates are of the 'C' on the + strand. numMethSamples, number of methylated samples. Percent methylation is averaged for both replicates and both strands (Methods), NA indicates no data. Originally generated by the ENCODE project (Varley, ENCODE submitted).

Supplementary Table S5. Sites tested in methylation analysis.
chrom, chromStart, chromEnd, Strand, the strand-oriented hg19 coordinates of 134 bp window around the motif. ID refers to the ChIP-seq peak; flag, nonzero indicates that the site was excluded from the regression. curIndivCpGs, locations of CpGs monitored by RRBS data. variableMe, TRUE indicates that the site exhibited variable methylation. sample.unmeth.size, sample.meth.size, the number of cell types whose binding sites are unmethylated and methylated, respectively. Slope, y.intercept, r.squared, p.value, and q.value refer to the regression.

Supplementary Table S6. Gene expression data for 19 cell types.
Results of RMA analysis.