Ad1_noMX:	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG
Ad2.1_TAAGGCGA	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGT
Ad2.2_CGTACTAG	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT
Ad2.3_AGGCAGAA	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT
Ad2.4_TCCTGAGC	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT
Ad2.5_GGACTCCT	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGT
Ad2.6_TAGGCATG	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGT
Ad2.7_CTCTCTAC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGT
Ad2.8_CAGAGAGG	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGT
Ad2.9_GCTACGCT	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTCTCGTGGGCTCGGAGATGT
Ad2.10_CGAGGCTG	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGGAGATGT
Ad2.11_AAGAGGCA	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGGAGATGT
Ad2.12_GTAGAGGA	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGGAGATGT
Ad2.13_GTCGTGAT	CAAGCAGAAGACGGCATACGAGATATCACGACGTCTCGTGGGCTCGGAGATGT
Ad2.14_ACCACTGT	CAAGCAGAAGACGGCATACGAGATACAGTGGTGTCTCGTGGGCTCGGAGATGT
Ad2.15_TGGATCTG	CAAGCAGAAGACGGCATACGAGATCAGATCCAGTCTCGTGGGCTCGGAGATGT
Ad2.16_CCGTTTGT	CAAGCAGAAGACGGCATACGAGATACAAACGGGTCTCGTGGGCTCGGAGATGT
Ad2.17_TGCTGGGT	CAAGCAGAAGACGGCATACGAGATACCCAGCAGTCTCGTGGGCTCGGAGATGT
Ad2.18_GAGGGGTT	CAAGCAGAAGACGGCATACGAGATAACCCCTCGTCTCGTGGGCTCGGAGATGT
Ad2.19_AGGTTGGG	CAAGCAGAAGACGGCATACGAGATCCCAACCTGTCTCGTGGGCTCGGAGATGT
Ad2.20_GTGTGGTG	CAAGCAGAAGACGGCATACGAGATCACCACACGTCTCGTGGGCTCGGAGATGT
Ad2.21_TGGGTTTC	CAAGCAGAAGACGGCATACGAGATGAAACCCAGTCTCGTGGGCTCGGAGATGT
Ad2.22_TGGTCACA	CAAGCAGAAGACGGCATACGAGATTGTGACCAGTCTCGTGGGCTCGGAGATGT
Ad2.23_TTGACCCT	CAAGCAGAAGACGGCATACGAGATAGGGTCAAGTCTCGTGGGCTCGGAGATGT
Ad2.24_CCACTCCT	CAAGCAGAAGACGGCATACGAGATAGGAGTGGGTCTCGTGGGCTCGGAGATGT

Supplementary Table 1: Oligo designs. A list of ATAC-seq oligos used for PCR.

Data table:

Name	DataSet
C-FOS	CFOSSTD
IRF3	IRF3IGGMUS
NFYA	NFYAIGGMUS
ZNF143	ZNF143166181APSTD
RAD21	RAD21IGGRAB
SMC3	SMC3AB9263IGGMUS
CTCF	CTCFSC15914C20STD
NFYB	NFYBIGGMUS
JUND Ab1	JUNDSTD
NFE2	NFE2SC22827STD
JUND Ab2	JUNDIGGRAB
P300B	P300BSTD
SREBP2	SREBP2IGGRAB
NRF1	NRF1IGGMUS
RFX5	RFX5200401194IGGMUS
E2F4	E2F4IGGMUS
ELK1	ELK112771IGGMUS
P300 Ab1	P300IGGMUS
P300 Ab2	P300SC584IGGMUS
MAX Ab1	MAXSTD
MAZ	MAZAB85725IGGMUS
BHLHE40	BHLHE40CIGGMUS
USF2	USF2IGGMUS
CHD2	CHD2AB68301IGGMUS
COREST	CORESTSC30189IGGMUS
STAT3	STAT3IGGMUS
TBLR1	TBLR1AB24550IGGMUS
SPT20	SPT20STD
STAT1	STAT1STD
TR4	TR4STD
TBP	TBPIGGMUS
MXI1	MXI1IGGMUS
ERRA	ERRAIGGRAB
EBF1	EBF1SC137065STD
MAX Ab2	MAXIGGMUS
MAFK	MAFKIGGMUS
SREBP1	SREBP1IGGRAB
CDP	CDPSC6327IGGMUS
IKZF1	IKZF1IKNUCLASTD
BRCA1	BRCA1A300IGGMUS
YY1	YY1STD
POL2 Ab1	POL2STD
POL2 Ab2	POL2IGGMUS
POL2s2	POL2S2IGGMUS
GCN5	GCN5STD
ZZZ3	ZZZ3STD
ZNF384	ZNF384HPA004051IGGMU
WHIP	WHIPIGGMUS
SIN3A	SIN3ANB6001263IGGMUS
CHD1	CHD1A301218AIGGMUS

Supplementary Table 2: ENCODE ChIP-seq data list. A list of data used for Figure 5d in the main text. All ChIP-seq data was downloaded from the Stanford/Yale/USC/Harvard (SYDH) ENCODE data repository available at the UCSC genome browser.

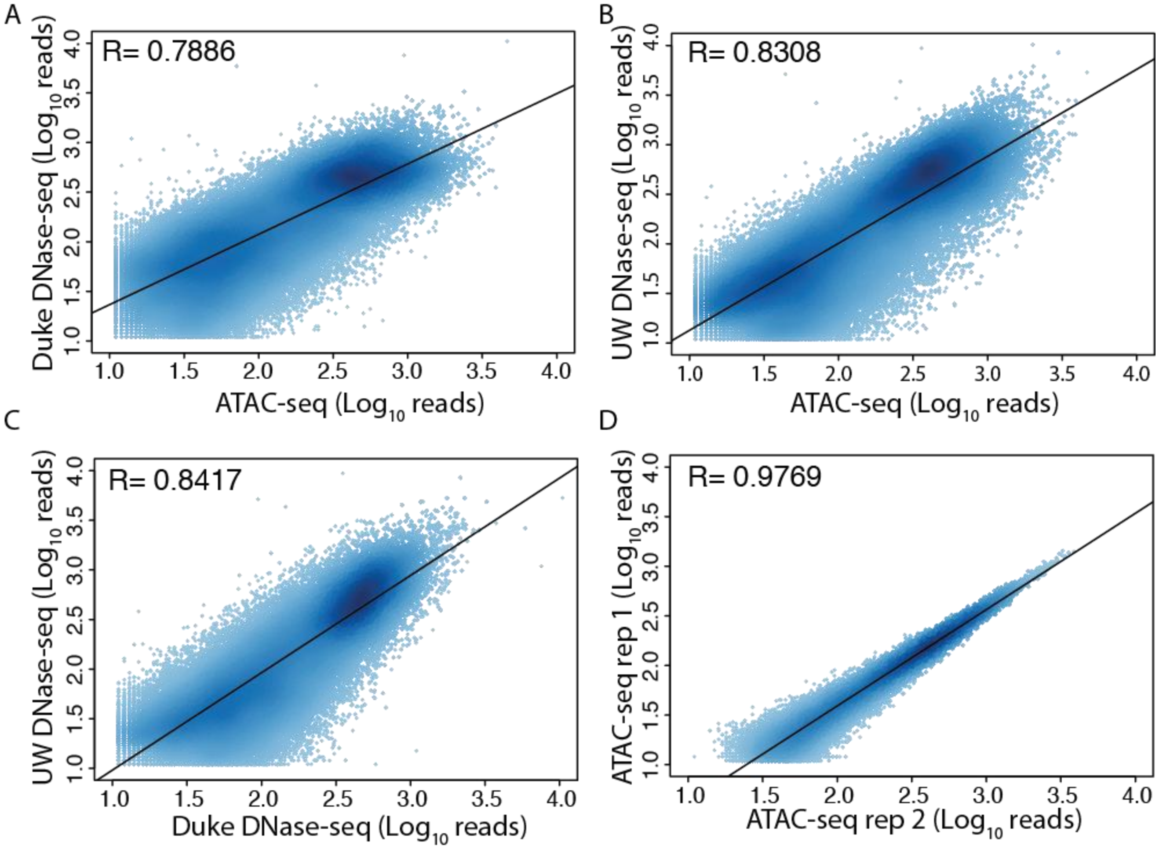
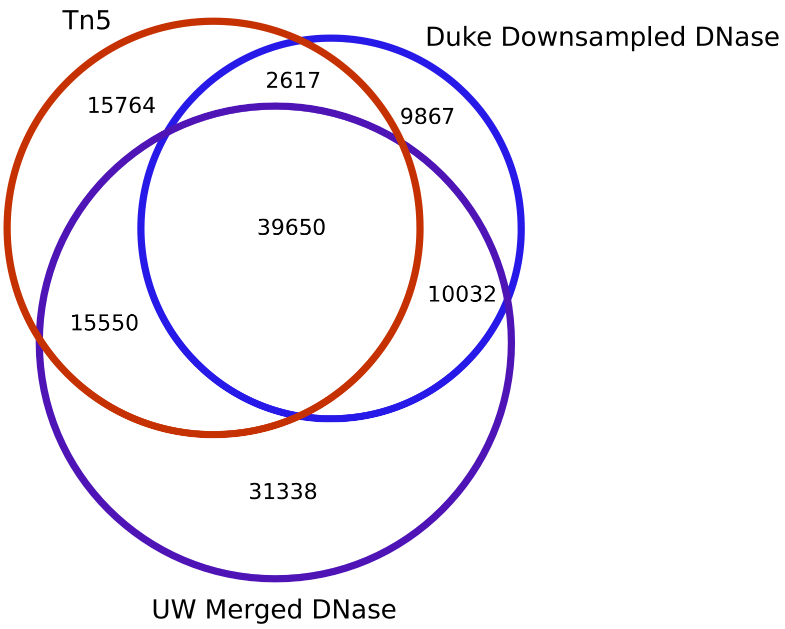


Figure S1: ATAC-seq peak intensity correlates well with DNase-seq peak intensity. Peaks in Duke DNase-seq (down sampled to 60 x 10^6 reads), UW DNase-seq (40 x 10^6 reads), and ATAC-seq data (60 x 10^6 paired-end reads) were called using ZINBA[1]. Because each data set has different read lengths we chose to filter for peaks within mappable regions (Duke DNase-seq = 20 bp reads, UW DNase-Seq = 36bp reads and ATAC-Seq = paired-end 50 bp reads). The log10(read intensity) was compared for (A) Duke DNase-seq and ATAC-seq, (B) UW DNase-seq and ATAC-seq, and (C) UW DNAse-seq and Duke DNase-seq. Technical reproducibility of ATAC-seq data is shown in D.



	ATAC Unique	UW Unique	Duke Unique	ATAC and UW	ATAC and Duke	UW and Duke	Intersect
ATAC-Seq	6.30%	4.36%	1.08%	10.83%	1.53%	2.09%	73.80%
UW DNase	1.32%	8.13%	0.68%	7.57%	0.33%	4.34%	77.62%
Duke DNase	2.10%	5.27%	14.80%	4.02%	1.44%	6.91%	65.46%

Figure S2: ATAC-seq captures a large fraction of DNase identified peaks. Peaks were called for all data sets using ZINBA (see **Fig. S1**). The venn-diagram shows overlap of the peak calls between each method. Below: The majority of ATAC-seq reads are in intense peaks that intersect with Duke and UW DNase-seq peaks. The total fraction of reads within peaks called from ATAC-seq, UW DNase-seq, and Duke DNase-seq, as well as the intersections of these data are shown. More than 65% of reads from all three methods are found in the intersection of the three methods' peaks, suggesting that strong well-stereotyped peaks are detected by all methods. Table cell color is proportional to fraction of reads.

Nature Methods: doi:10.1038/nmeth.2688

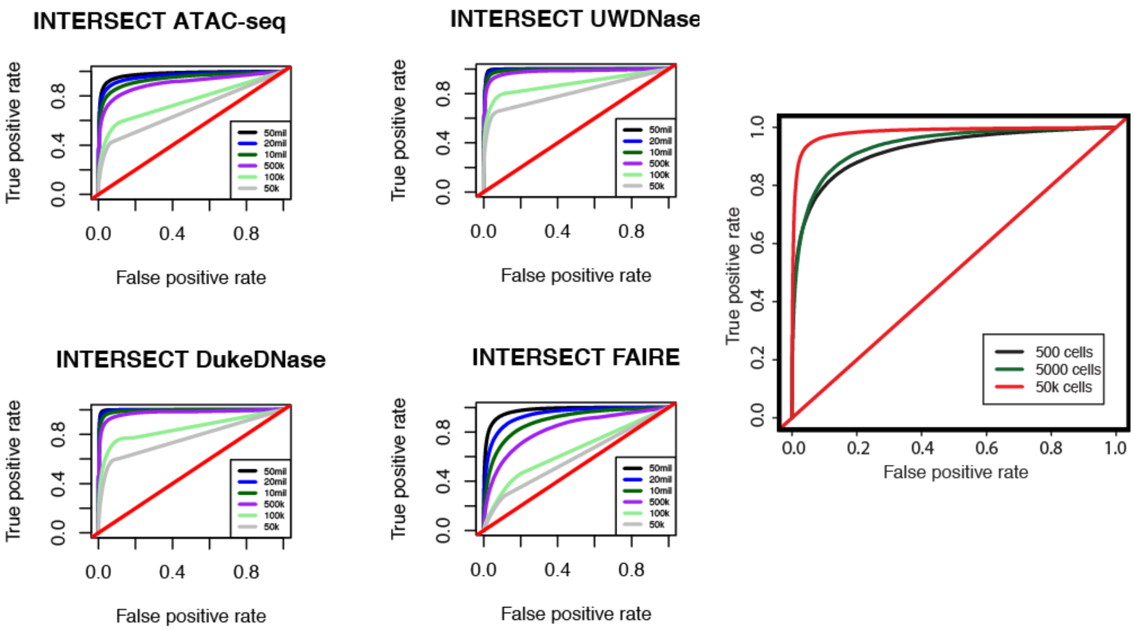


Figure S3: Performance of ATAC-seq in GM12878 cells by read and cell number. For each analysis the number of reads overlapping the set of open chromatin regions identified by Duke DNase, UW DNase and FAIRE in GM12878 cells were compared to a set of background regions. To determine the read depth required for detecting open chromatin sites sensitivity and specificity was assessed at varying read depths, including 50k, 100k, 500k, 10 million, 20 million and 50 million reads (left four panels). Performance of ATAC-seq in GM12878 cells was assessed using 500, 5,000 or 50,000 cells as starting material (right panel).

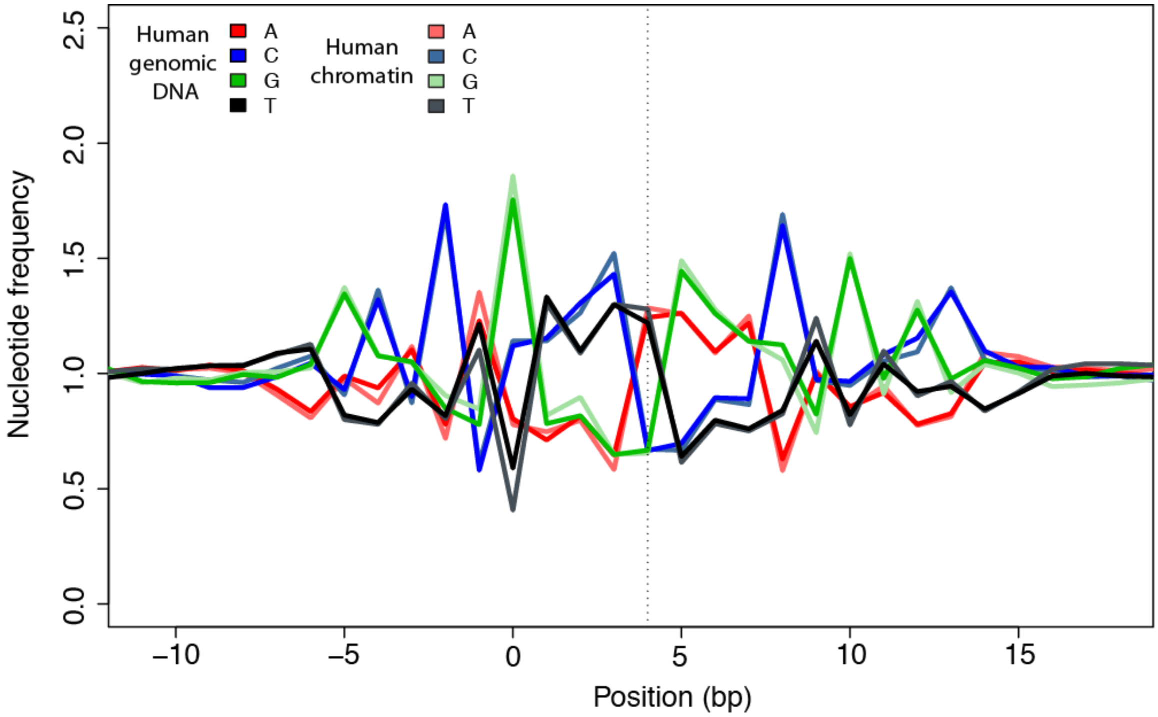


Figure S4: Tn5 insertion preferences in genomic DNA and chromatin. Nucleotide frequency scores represent the observed nucleotide frequency of each base, nucleotide frequencies are normalized to 1. The x=0 position represents the read start, and the dotted line represents the symmetry axis of the Tn5 dimer. We see no substantial differences between Tn5 insertion preferences between purified genomic DNA and human chromatin, suggesting that the local insertion preference into chromatin is identical to that found in naked genomic DNA. These reported sequence preferences are similar to those previously reported (main text ref. 11).

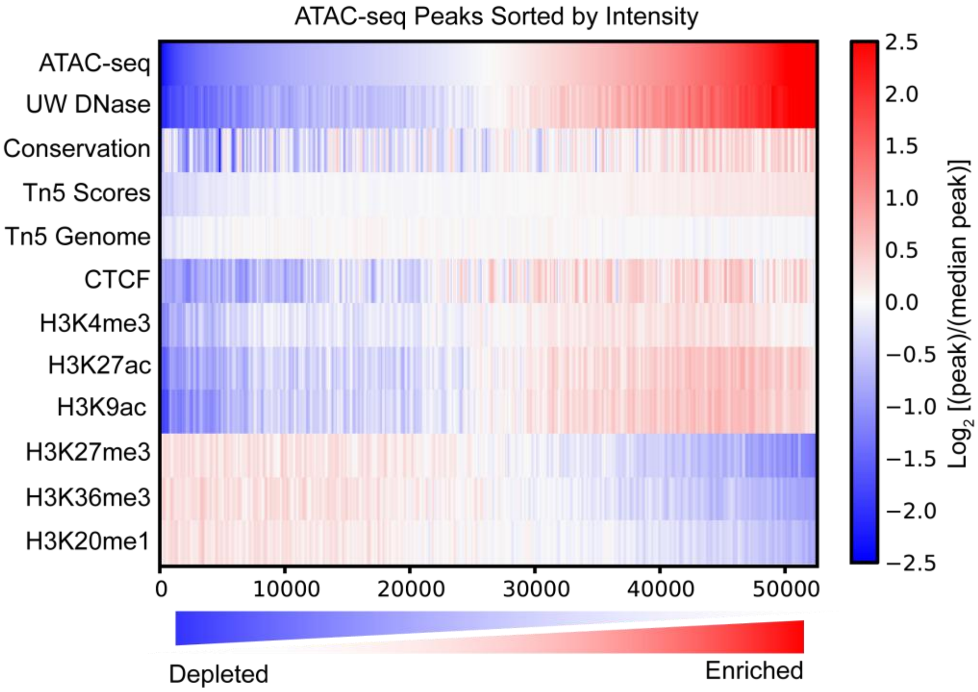


Figure S5. Correlation of ATAC-seq peak intensity with various features of the genome. The following data represents the average intensity per base of each feature at every ATAC-seq peak. All ENCODE ChIP data was normalized to input, data has been processed using a sliding window of 200 peaks. We observed that ATAC-seq peak intensity was most strongly correlated with DNase hypersensitivity (see **Fig. S1** for more information) and CTCF (see **Fig. 3** in the main text). ATAC-Seq is only moderately correlated with Tn5 sequence preference. Tn5 scores represents a theoretical insertion bias derived using the PWM (see **Fig. S4**), Tn5 genome is derived from data produced from Adey et al. (main text ref. 11). Briefly, data was trimmed and processed identically to what was described in the analysis. We also saw that ATAC-seq was correlated with histone marks associated with active chromatin (H3K4me3, H4K27ac and H3k9ac), and anti-correlated with histone marks associated with inactive chromatin (H3K27me3) and gene bodies (H3K36me3 and H4K20me1). We also note a moderate enrichment for conserved bases at higher intensity peaks.

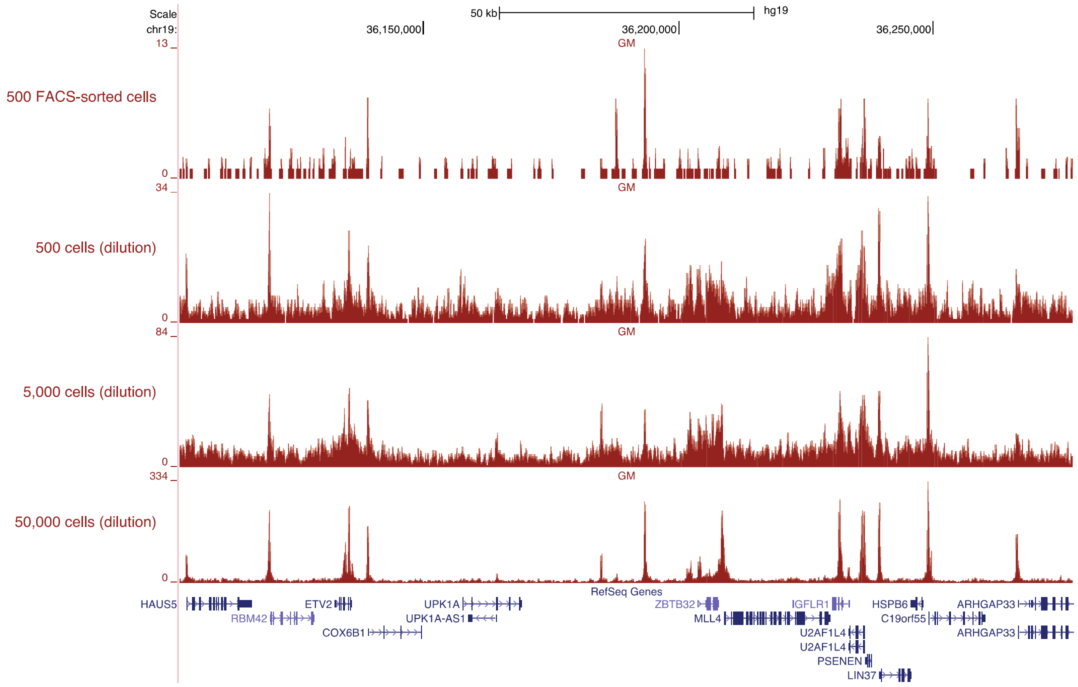


Figure S6: ATAC-seq of various cell numbers. A representative UCSC genome browser track of data from different starting numbers of cells for ATAC-seq. This same locus is also shown in **Fig. 1b** of the main text. In order: 500 cells were isolated using FACS and two replicates of 500 cells and 5,000 cells were done by a simple dilution from cell culture. For comparison, the bottom track represents 50,000 cells, also show in **Fig. 1b**. This figure demonstrates that we are able to capture open chromatin sites from as few as 500 cells.

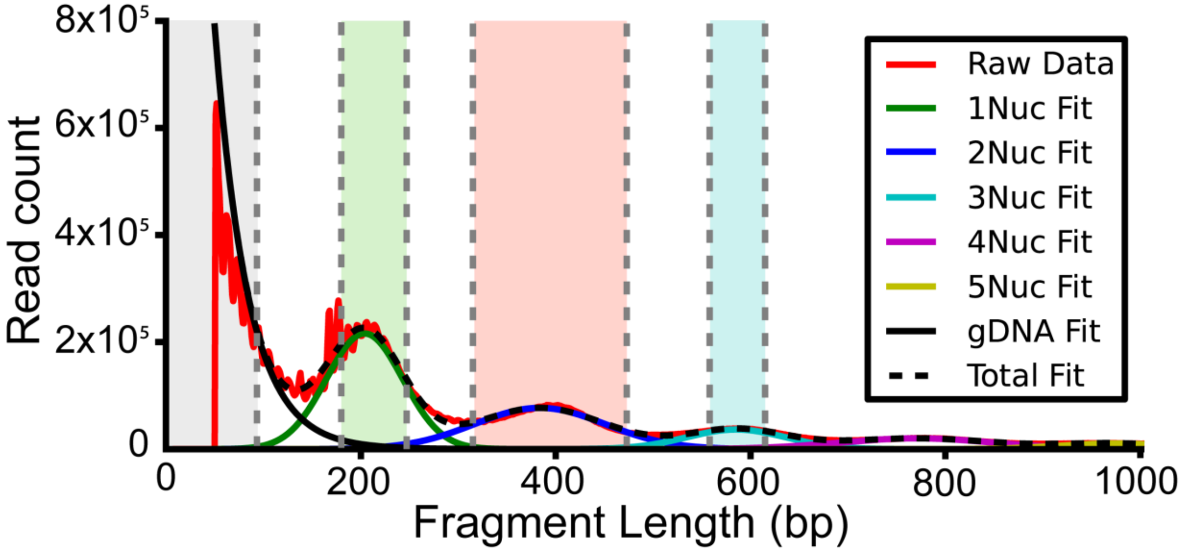


Figure S7: Fitting nucleosome peaks in ATAC-seq fragment size distribution to enable nucleosome occupancy measurements. The observed fragment distribution was partitioned into four populations of reads – reads expected to originate from open DNA, and reads that span 1, 2 or 3 putative nucleosomes. To enable this partitioning of the data, the ATAC-seq fragment distribution was fit to the sum of 1) an exponential function for fragment distribution pattern at insert sizes below one nucleosome, and 2) 5 Gaussians to the distributions arising from protection from one, two, three, four and five nucleosomes. The sum of these fits is shown (black dotted line) is similar to the observed fragment distribution (blue line). Vertical dotted lines are boundaries for identification of fragments as originating from the nucleosome-free (<100bps), 1-nucleosome, 2-nucleosome and 3-nucleosome regions. Dotted lines were set to ensure that <10% of fragments originate from neighboring, as defined by our fit.

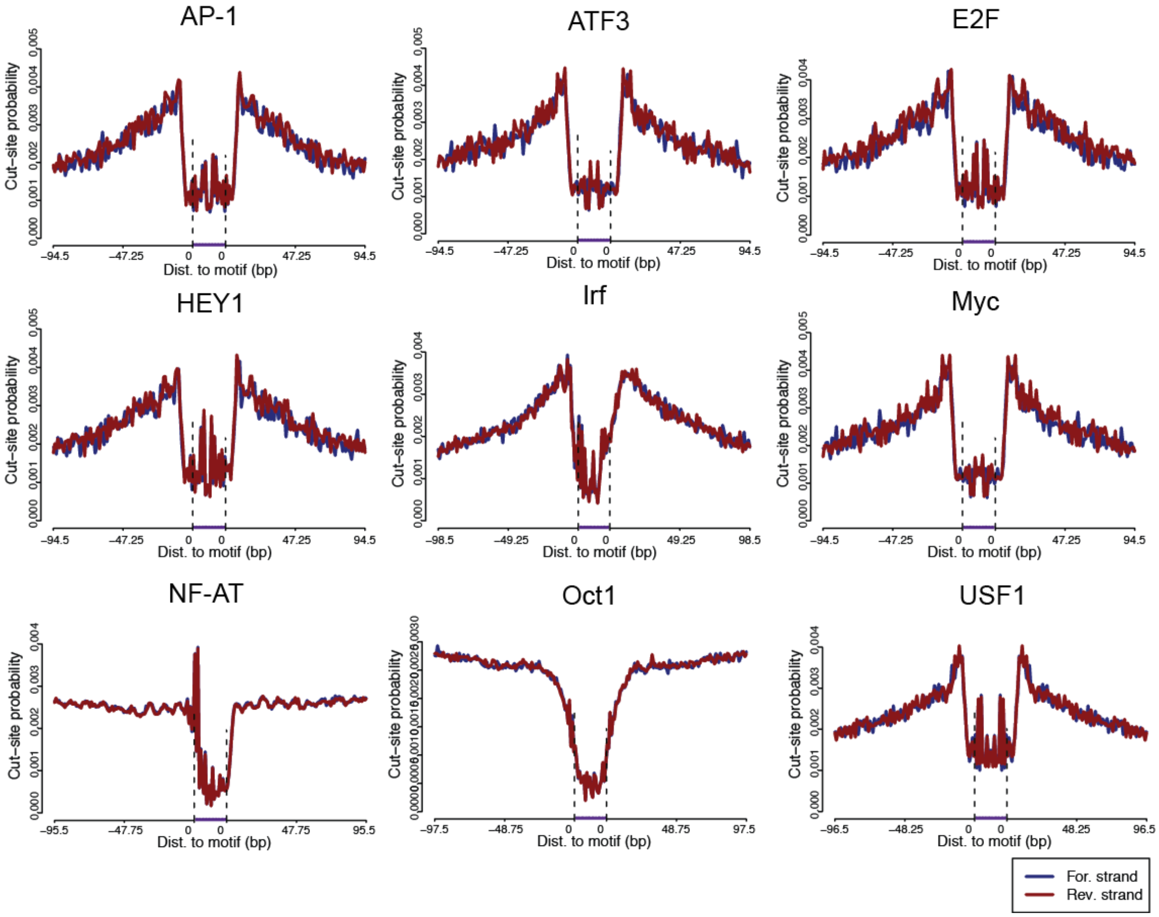


Figure S8: Select set of transcription factor footprints detected by ATAC-seq in GM12878 cells. For the indicated transcription factors the aggregate signal of ATAC-seq reads were computed using CENTIPEDE on the genome-wide sets of sites matching the corresponding motif. Reads were calculated in the region +/-100 bp of the motif boundary. The vertical dashed lines indicate the boundaries of the motifs.

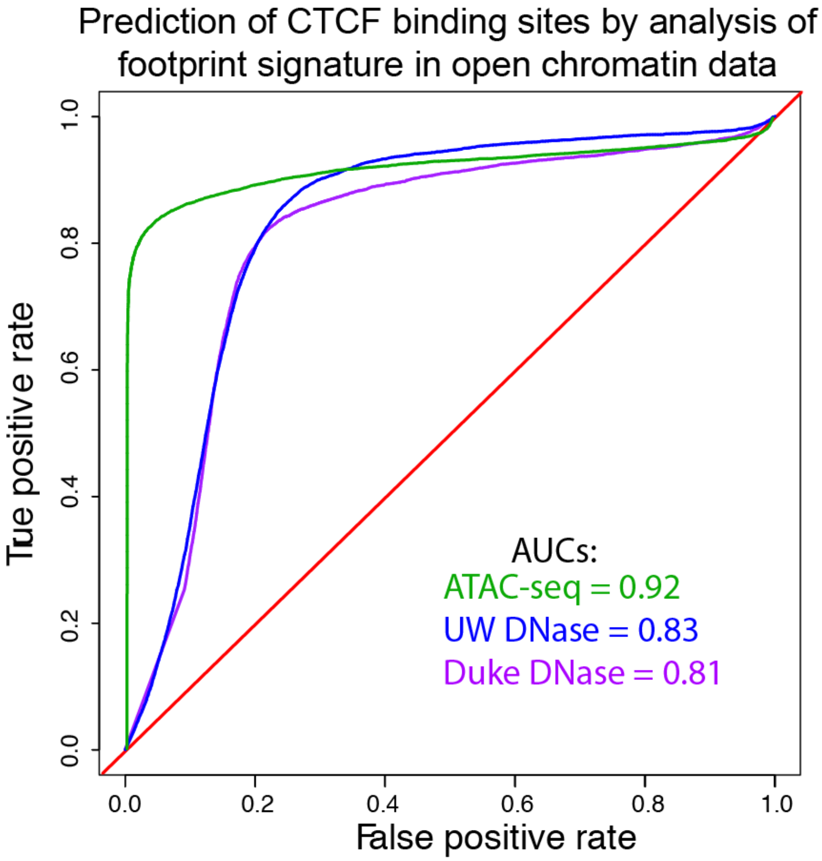


Figure S9: Prediction of CTCF binding sites using ATAC-seq and DNase footprinting with CENTIPEDE. Prediction of CTCF binding sites was assessed using the genome-wide set of CTCF motifs sorted by the posterior probability reported by CENTIPEDE. Those overlapping CTCF ChIP-seq peaks were used as the positive set and all others were considered as the negative set. This yielded an area under the curve (AUC) of 0.92, which suggests specific and sensitive binding inference for CTCF. Duke DNase and UW DNase data were used with the same settings of CENTIPEDE, and ROC plots are shown. ATAC-seq data consisted of 198 x 10^6 paired reads, Duke DNase-comprised 245 x 10^6 reads, and UW DNase comprised 48 x 10^6 reads.

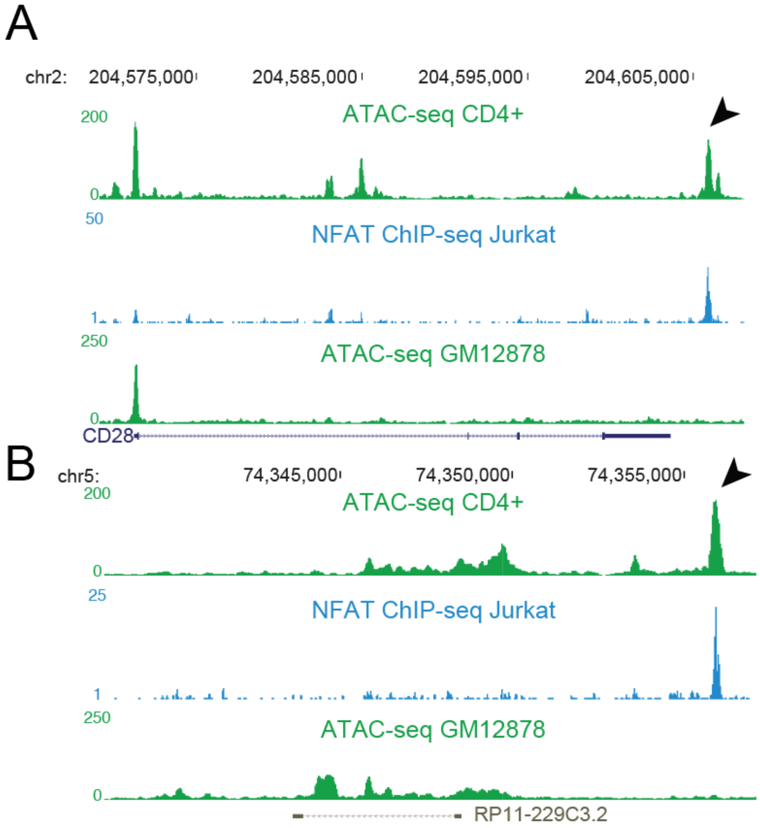


Figure S10: T-cell specific NFAT regulation: Examples of T-cell-specific NFAT target genes predicted by ATAC-seq and confirmed by alignment with NFAT ChIP-seq (data from main text ref 35).

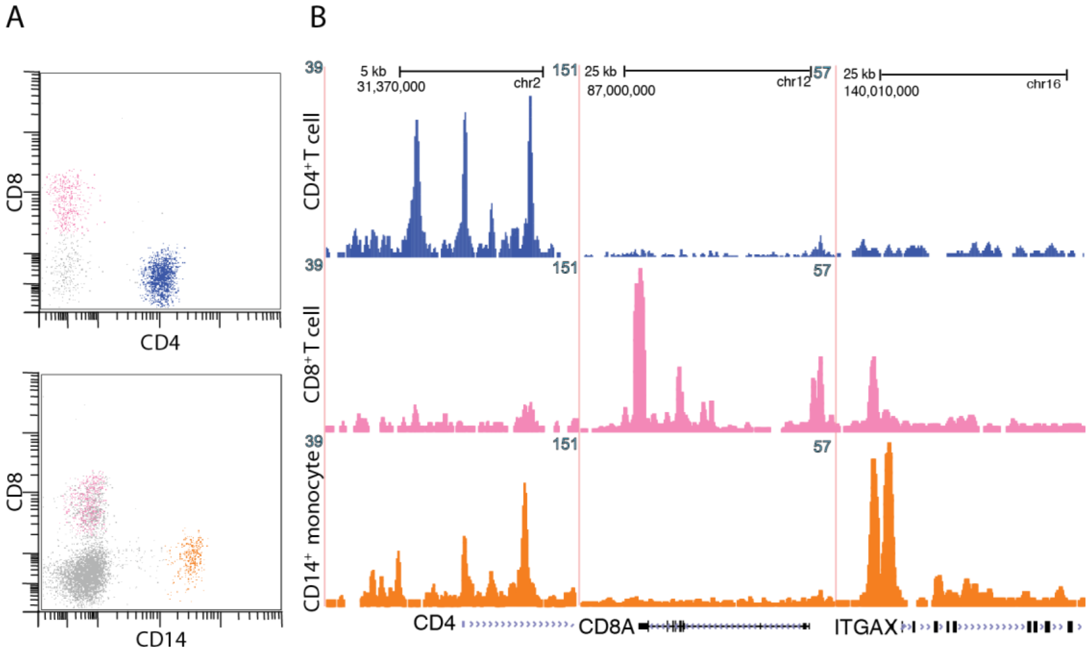


Figure S11: ATAC-seq of FACS-purified cell populations from human blood. (A) From a standard blood draw, we used Fluorescence-Activated Cell Sorting (FACS) to purify CD4+ T-cells, CD8+ T-cells, and CD14+ monocytes. Each population generated successful ATAC-seq data (B) and revealed cell-type specific open chromatin sites at known lineage-specific genes.

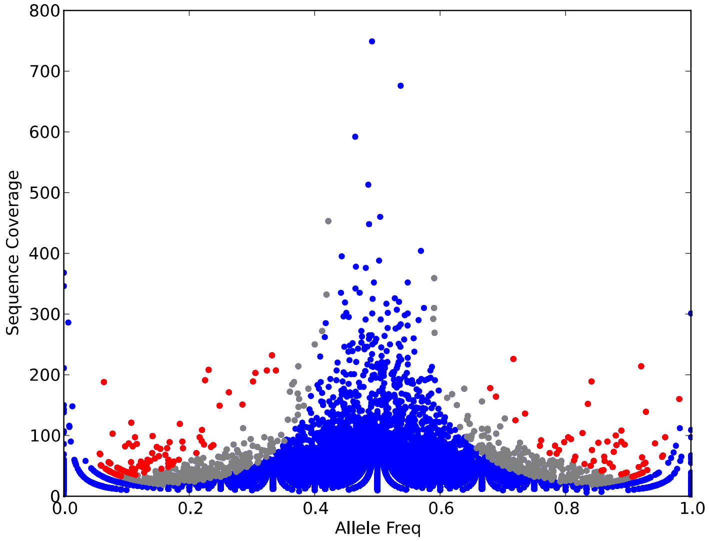


Figure S12: Detection of allele specific open chromatin in GM12878 cells with ATAC-seq. Using publicly available variant data, we measured the allele frequency in open chromatin regions at putative heterozygous loci. Because of potential for spurious heterozygous sites, we required more than two reads to validate the heterozygosity of the allele. Red points (n=167) are candidate allele specific open chromatin sites at $p<10^{-5}$, while grey (n=900) represent candidates at p<0.01. P-values were calculated using a Bayesian model developed by Audic et al.[2].

Supplemental References:

1. Rashid, N. U., Giresi, P. G., Ibrahim, J. G., Sun, W. & Lieb, J. D. ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions. *Genome Biol* **12,** R67 (2011).
2. Audic, S. & Claverie, J. M. The significance of digital gene expression profiles. *Genome Research* **7,** 986–995 (1997).