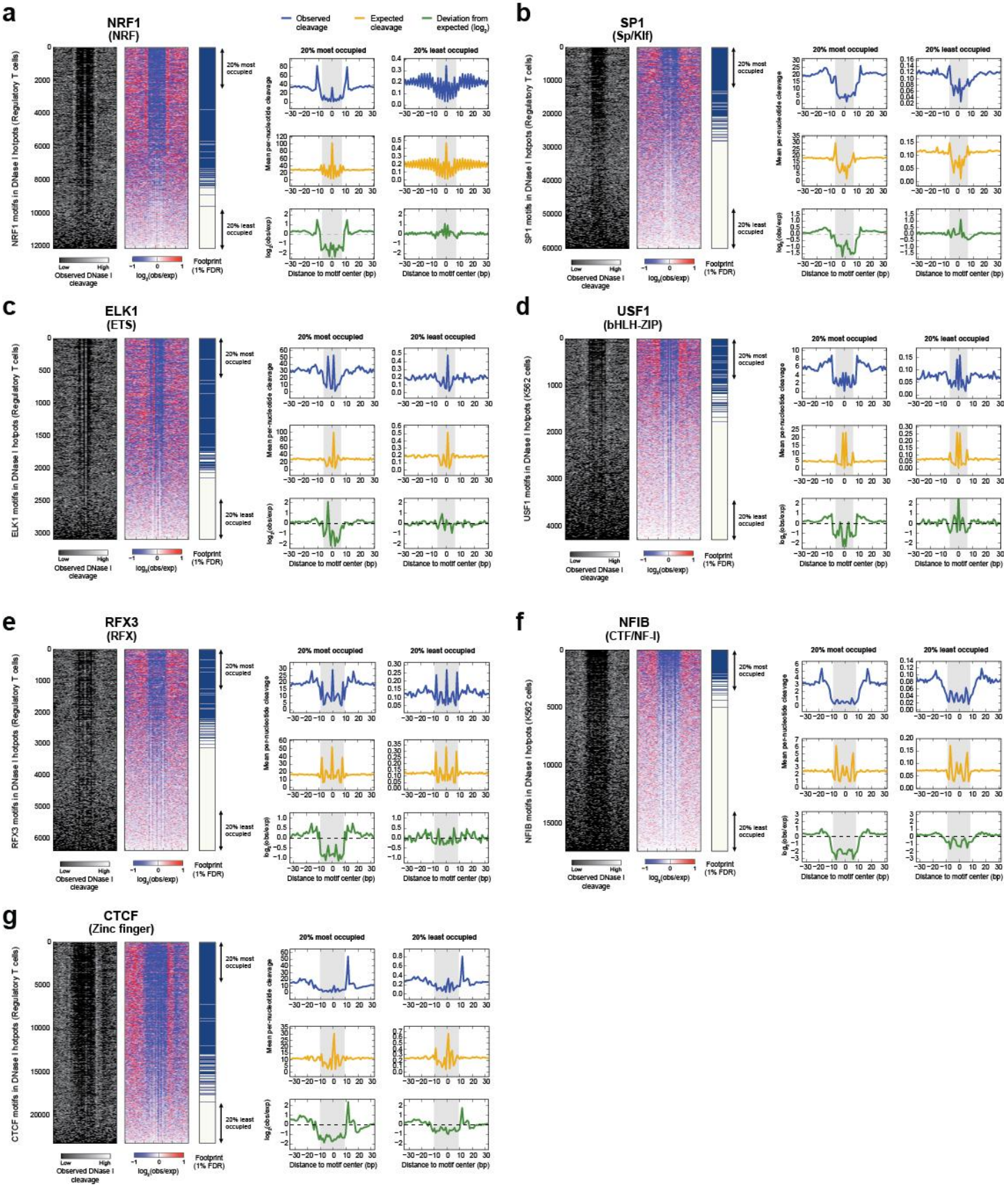
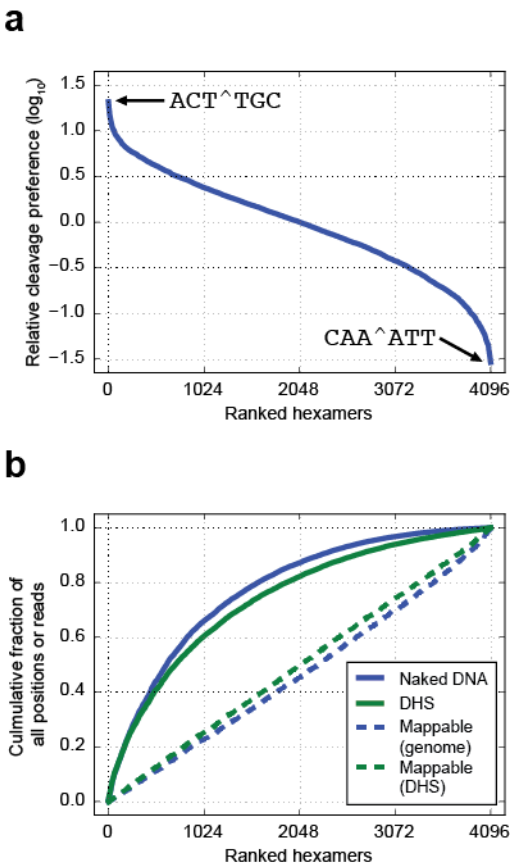


Nature Methods: doi:10.1038/nmeth.3768

Supplementary Figure 1

Aggregated DNase I cleavage patterns for TF recognition sequences reflecting diverse DNA-binding domains.

(a) Heatmaps of per-nucleotide DNase I cleavages and discovered footprints surrounding NRF1 recognition sequences. Left, observed cleavages. Right, the ratio of the observed cleavages to expected cleavages computed by reassigning tags to a hexamer model DNase I cleavage bias. Blue ticks indicate that the recognition sequence has an associated DNase I footprint. Line plots show the aggregate profile of mean per-nucleotide DNase I cleavages at the 20% most (left column) and 20% least (right column) accessible NRF1 recognition sequences. Top row, observed cleavages. Middle, expected cleavages computed using the hexamer model. Bottom, the $\log_2$ ratio of observed to expected. (b-g) The same as (a) for the recognition sequences for (b) SP1, (c) ELK1, (d) USF1, (e) RFX3, (f) NFIB, and (g) CTCF within accessible chromatin. In each case the cleavage patterns at occupied templates (coinciding with *de novo* TF footprint calls) parallel known structural features of the respective DNA binding domains.



Supplementary Figure 2

General features of DNase I sequence preference.

(a) Relative cleavage preference of all 4,096 hexamers with respect to the median hexamer as determined by deep sequencing (~100 million tags) of a DNase I digestion of deproteinized DNA from human IMR90 cells (data from ref. 4). (b) Biased hexamers contribute disproportionately to total DNase I cleavages for both naked DNA and chromatin (regulatory T cells cleavages mapping within DHS) when compared to the 36 bp mappable genome. Shown is the cumulative fraction all mappable positions or sequencing tags within respect to their hexamer context. Hexamers are ranked by decreasing cleavage preference as in a.

Supplementary Box 1. Joint modeling nuclease cleavage action and TF kinetics

As originally described by Galas and Schmitz[1], the relation between factor occupancy and nuclease cleavage can be described by the following sets of equations. *P* represents the concentration of transcription factor, B_i represents a single nucleotide *i* in genome, and PB_i represents the nucleotide *i* is bound by the transcription factor.

$$\mathrm{P} + \mathrm{B}_i \underset{k_{i,-1}}{\overset{k_{i,1}}{\rightleftharpoons}} \mathrm{PB}_i$$

Assuming binding reaction is rapid equilibrium, then $K_{i,d}$ represents the dissociation constant and the concentration of transcription factor bound nucleotide *i* can be defined by the follow set of equations:

$$[PB_i] = \frac{k_{i,1}}{k_{i,-1}}[P][B_i] = K_{i,D}[P][B_i]$$

The total amount of nucleotide ($[B_i]_t$) is the sum of the bound and unbound, which can be defined in terms of the equilibrium equation derived above. And now the amount of unbound nucleotide can be defined by the following relationship parameterized by dissociation constant for the transcription factor.

$$[B_i]_t = [B_i] + [PB_i]$$

$$[B_i] = \frac{[B_i]_t}{1 + K_{i,D}[P]}$$

DNase I competes with the transcription factor for access to the nucleotide, and as such the rate of cleavage is related to the occupancy of the transcription factor:

$$\mathrm{D} + \mathrm{B}_i \xrightarrow{k_{i,S}} \mathrm{D} + \mathrm{B}_i^{\circ}$$

$$\frac{d[B_i^{\circ}]}{dt} = k_{i,S}[D][B_i] = k_{i,S}[D]\frac{[B_i]_t}{1 + K_{i,D}[P]}$$

Thus, the rate of cleavage products at a nucleotide (B_i°) is related to the occupancy of a transcription factor ($K_{i,D}$) and the sequence preference of DNase I ($k_{i,S}$). As such, the occupancy of transcription factors can be determined by comparing the rates of cleavage at a specific (i.e., recognition element) vs. a non-specific site (i.e., adjacent nucleotide).

$$\frac{[B_i]}{[B_j]} \sim \frac{\frac{d[B_i^\circ]}{dt}}{\frac{d[B_j^\circ]}{dt}}$$

Which can be generalized to a model that is dependent on four parameters:

$$\frac{[B_i]}{[B_j]} \sim \left[\frac{k_{i,S}}{k_{j,S}}\right] \frac{1 + K_{j,D}[P]}{1 + K_{i,D}[P]}$$

The parameters $K_{i,d}$ and $K_{j,d}$, are the binding constants for nucleotide i and j respectively. Assuming that one of the sites (*i*) corresponds to a preferred or specific recognition element, while the other (*j*) is a non-specific site, then it is expected that $K_{i,D} >> K_{j,D}$ which would result in large differences in observed cleavages at these two nucleotides. At sites with low occupancy ($K_{i,D} \approx K_{j,D}$), the observed differences cleavage rates reflect the inherit sequence prefence of DNase I itself (parameterized by $K_{i,S}$ and $K_{j,S}$).

The biophysical properties described above provide the basis for kinetic profiling of TF occupancy, which has been widely exploited to analyze individual templates and TFs (reviewed in refs. 2,3).

Supplementary Table 1. Experimental considerations for digital genomic footprinting

Parameter	Optimum	Common pitfalls
Fragment size	Majority of fragments <125 bp. *Fragments in this size range derive chiefly from cleavages spanning TF occupancy sites within a single regulatory DNA template (ref. 5).*	Poor isolation of small fragments, contamination with nucleosomal derived fragments
Signal-to-noise ratio (SNR) *(Critical parameter)*	Signal proportion of tags (SPOT) score > 0.35, optimally >0.5. *Samples with low SPOT scores signal an intrinsic defect that cannot be 'rescued' by deeper sequencing.*	Poor cell/nuclear condition; over-digestion of chromatin resulting in nucleosomal derived fragments; poor enrichment of small fragments
Library complexity and sequencing depth	>200 million unique fragments for large-scale de novo footprint detection. *Lower read depths may be sufficient for quantifying occupancy of individual motif instances*.	Over amplification, library selection
Mappability to reference	>75% total uniquely mapping reads to nuclear genome reference (mouse/human)	Mitochondrial and/or adapter sequence contamination
Duplicate reads (resulting from amplification during library construction)	<1% of uniquely mapping reads	Over amplification, poor library complexity

Supplementary References

1. Galas, D. J. & Schmitz, A. DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res.* **5,** 3157–3170 (1978).
2. Dabrowiak, J. C., Goodisman, J. & Ward, B. Quantitative DNA footprinting. *Methods Mol. Biol.* **90,** 23–42 (1997).
3. Hampshire, A. J., Rusling, D. A., Broughton-Head, V. J. & Fox, K. R. Footprinting: a method for determining the sequence selectivity, affinity and kinetics of DNA-binding ligands. *Methods* **42,** 128–140 (2007).
4. Lazarovici, A. *et al.* Probing DNA shape and methylation state on a genomic scale with DNase I. *Proc. Natl. Acad. Sci. USA.* **110,** 6376–6381 (2013).
5. Vierstra, J., Wang, H., John, S., Sandstrom, R. & Stamatoyannopoulos, J. A. Coupling transcription factor occupancy to nucleosome architecture with DNase-FLASH. *Nat. Meth.* **11,** 66–72 (2014).