Published in final edited form as:

Nature. ; 474(7352): 516–520. doi:10.1038/nature10002.

Determinants of nucleosome organization in primary human cells

Anton Valouev[1], **Steven M. Johnson**[2], **Scott D. Boyd**[1], **Cheryl L. Smith**[3], **Andrew Z. Fire**[1,3], and **Arend Sidow**[1,3]

[1]Department of Pathology, Stanford University School of Medicine, 300 Pasteur Dr., Stanford, CA, 94305

[2]Department of Microbiology and Molecular Biology, Brigham Young University, 757 WIDB, Provo, UT, 84602-5253

[3]Department of Genetics, Stanford University School of Medicine, Pasteur Dr., Stanford, CA, 94305

Abstract

Nucleosomes are the basic packaging units of chromatin, modulating accessibility of regulatory proteins to DNA and thus influencing eukaryotic gene regulation. Elaborate chromatin remodeling mechanisms have evolved that govern nucleosome organization at promoters, regulatory elements, and other functional regions in the genome[1]. Analyses of chromatin landscape have uncovered a variety of mechanisms, including DNA sequence preferences, that can influence nucleosome positions[2–4]. To identify major determinants of nucleosome organization in the human genome, we utilized deep sequencing to map nucleosome positions in three primary human cell types and *in vitro*. A majority of the genome exhibited substantial flexibility of nucleosome positions while a small fraction showed reproducibly positioned nucleosomes. Certain sites that position *in vitro* can anchor the formation of nucleosomal arrays that have cell type-specific spacing *in vivo*. Our results unveil an interplay of sequence-based nucleosome preferences and non-nucleosomal factors in determining nucleosome organization within mammalian cells.

Previous studies in model organisms[3–7] as well as initial analyses in human cells[8] have identified fundamental aspects of nucleosome organization. We here focus on the dynamic relationships between sequence-based nucleosome preferences and chromatin regulatory function in primary human cells. We mapped tissue-specific and DNA-encoded nucleosome organization across granulocytes and two types of T-cells (CD4+ and CD8+) isolated from the blood of a single human donor, by isolating cellular chromatin and treating it with

Correspondence to: Andrew Z. Fire; Arend Sidow.

Author contributions

AV, SMJ, AS and AZF designed the experiments. SMJ, AV, CLS and SB performed the experiments. AV designed and carried out analyses with input from AS, AZF, and SMJ. AV, AS and AZF wrote the manuscript.

Data availability. All sequence data were submitted to Sequence Read Archive (accession number GSE25133). Sites containing strongly positioned *in vitro* nucleosomes are available as a supplementary data file.

micrococcal nuclease (MNase) followed by deep sequencing of the resulting nucleosome-protected fragments (Methods, Supplementary Fig. 1). To provide sufficient depth for both local and global analyses, we used high-throughput SOLiD technology, generating 584, 342, and 343 million mapped reads for granulocytes, CD4+, and CD8+ T-cells, respectively. These are equivalent to 16x–28x genome coverage by 147 bp nucleosome footprints (cores; see Methods). The depth of sequence was critical for our subsequent analysis: while shallower coverage can illuminate features of nucleosome positions through statistical analysis (e.g. 6,8), any definitive map and thus comparison of static and dynamic positioning requires high sequence coverage throughout the genome.

To provide complementary data on purely sequence-driven nucleosome positioning in the absence of cellular influences, we reconstituted genomic DNA *in vitro* with recombinantly derived histone octamers to produce *in vitro* nucleosomes (Methods, Supplementary Fig. 2), and generated over 669 million mapped reads, representing 32x core coverage of the genome. To identify primary nucleosome positioning sites in DNA, the reconstitution was performed under conditions of DNA excess (see methods). We also generated a control dataset of 321 million mapped reads from MNase-digested naked DNA (Supplemental Materials). In the population of granulocytes (our deepest *in vivo* data set), over 99.5 % of the mappable genome is engaged by nucleosomes (Methods), and 50 percent of nucleosome-depleted bases occur in regions shorter than 160 bp.

We first focused on global patterns of nucleosome positioning and spacing by calculating fragment distograms and phasograms[6,7,9]. Distograms (histograms of distances between mapped reads' start positions aligning in opposing orientation, Supplementary Fig. 3A) reveal the average core fragment size as a peak if there are many sites in the genome that contain consistently positioned nucleosomes. A positioning signal that is strongly amplified by conditioning the analysis on sites with 3 or more read starts (reflecting a positioning preference; 3-pile subset), is present not only *in vivo* (Fig. 1A), but also *in vitro* (Fig. 1B), demonstrating that many genomic sites bear intrinsic, sequence-driven, positioning signals. Phasograms (histograms of distances between mapped reads' start positions aligning in the same orientation, Supplementary Fig. 3B) reveal consistent spacing of positioned nucleosomes by exhibiting a wave-like pattern with a period that represents genome-average internucleosome spacing. In granulocytes, the wave peaks are 193 bp apart (Fig. 1C, adjusted R^2=1, p-value<10^{-15}), which, given a core fragment length of 147 bp, indicates an internucleosome linker length of 46 bps. By contrast, the phasograms of both types of T-cells have spacing that is wider by 10 bp (Fig. 1D), equivalent to a 56 bp average linker length. These results are consistent with classical observations of varying nucleosome phases in different cell types[10,11]. Linker length differences have been tied to differences in linker histone gene expression[12,13], which we found to be 2.4 times higher in T-cells compared to granulocytes (84 RPKM[14] vs.35 RPKM). The *in vitro* phasogram (Fig. 1E) reveals no detectable stereotypic spacing of positioned nucleosomes, demonstrating a lack of intrinsic phasing among DNA-encoded nucleosome positioning sites.

Using a positioning stringency metric (Methods; Supplementary Fig. 4) that quantifies the fraction of defined nucleosome positions within a given segment, we calculated the fraction of the genome that is occupied by preferentially positioned nucleosomes at different

stringency thresholds. The maximum number of sites at which some positioning preference can be detected statistically is 120 M, covering just over 20% of the genome (Supplementary Fig. 5)at the low stringency of 23%. Thus, the majority of nucleosome positioning preferences is weak, and nucleosomes across the majority of the human genome are not preferentially positioned, either by sequence or by cellular function.

We next focused on how transcription and chromatin functions affect nucleosome organization regionally. For each cell type, we generated deep RNA-seq data and binned genes into groups according to their expression levels. The average spacing of nucleosomes was greatest within silent genes (CD4+ T-cells, 206 bp, Fig. 2A) and decreased by as much as 11 bp as the expression levels went up (t-statistic p-value 6.5×10^{-34}). This suggests that transcription-induced cycles of nucleosome eviction and reoccupation cause denser packing of nucleosomes and slight reduction in nucleosome occupancy (Supplementary Fig. 6). On the basis of this result, we hypothesized that higher-order chromatin organization as implied by specific chromatin modifications might be associated with specific spacing patterns. Using previously published ChIP-seq data, we identified regions of enrichment[15] for histone modifications that are found within heterochromatin (H3K27me3, H3K9me3)[16], gene-body euchromatin (H4K20me1, H3K27me1)[16], or euchromatin associated with promoters and enhancers (H3K4me1, H3K27ac, H3K36ac)[17], and estimated spacing of nucleosomes for each of these epigenetic domains. We found that active promoter-associated domains contained the shortest spacing of 178–187 bp, followed by a larger spacing of 190–195 bp within the body of active genes, while heterochromatin spacing was largest at 205 bp (Fig. 2B). These results reveal striking heterogeneity in nucleosome organization across the genome that depends on global cellular identity, metabolic state, regional regulatory state, and local gene activity.

To characterize DNA signals responsible for consistent positioning of nucleosomes, we identified 0.3 million sites occupied *in vitro* by nucleosomes at high stringency (> 0.5; Methods). The region occupied by the center of the nucleosome (dyad) exhibits a significant increase in G/Cusage(Poisson p-value $< 10^{-100}$; Fig. 3A). Flanking regions increase in A/T usage as the positioning strength increases (Fig. 3B). A subset of *in vitro* positioned nucleosomes (stringency > 0.5) which are also strongly positioned *in vivo* (stringency > 0.4) revealed increased A/T usage within the flanks (Fig. 3C) compared to *in vitro*-only positioning sites (Fig. 3A), which underscores the importance of flanking repelling elements for positioning *in vivo*. We term such elements with strong G/C cores and A/T flanks “container sites” to emphasize the proposed positioning mechanism (Fig. 3D). This positioning signal is different from a 10 bp dinucleotide periodicity observed in populations of nucleosome core segments isolated from a variety of species[19,20] and proposed to contribute to precise positioning and/or rotational setting of DNA on nucleosomes[20] on a fine scale(Supplementary Fig. 7). G/C rich signals are known to promote nucleosome occupancy[18,21], while AA-rich sequences repel nucleosomes[4], and our data demonstrate that precise arrangement of a core-length attractive segment flanked by repelling sequences can produce a strongly positioned nucleosome (Fig. 3D).

Dyad frequencies around container sites (Fig 3E) show a strong peak of enrichment *in vivo*, confirming that DNA positions nucleosomes *in vivo* over these sites. Additionally, wave-like

patterns emanate from these sites *in vivo* (but not *in vitro*), reflecting the nucleation of phased arrays by positioned cellular nucleosomes. Viewing these results in light of the nucleosome barrier model[22], which proposes that nucleosomes are packed into positioned and phased arrays against a chromatin barrier, we conclude that sequence-positioned nucleosome can initiate propagation of adjacent stereotypically positioned nucleosomes. Importantly, wave periods around container sites are shorter in granulocytes than in T-cells, allowing tissue-specific variation in linker length (Fig. 1D) to alter placement of nucleosomes over distances of as much as 1 kb from an initial container site. Functional consequences of such rearrangements might include global shifts in regulatory properties that could contribute to distinct transcription factor accessibility profiles in different cell types.

The cellular environment can drive nucleosomes to sequences not intrinsically favorable to being occupied, as evident in a genome-wide comparison of observed nucleosome coverage of all possible tetranucleotides between the granulocyte and the *in vitro* data (Fig. 4A). *In vitro*, nucleosome occupancy is strongly associated with AT/GC content, but this preference is abolished *in vivo*; the exception are C/G rich tetramers that contain CpG dinucleotides, which show a 30% reduction in apparent nucleosome occupancy despite having high core coverage *in vitro*. Consistent with this, CpG islands are five-fold depleted for observed nucleosome coverage *in vivo* (Fig. 4B). No such decrease is observed in the *in vitro* dataset.

We hypothesize that the decreased nucleosome occupancy of promoters could be due to promoter-related functions of mammalian CpG islands, similar to promoter-associated nucleosome-free regions observed in flies[23] and yeast[5], which do not have CpG islands. We therefore analyzed transcription-dependent nucleosome packaging around promoters. As in other organisms[23–27], promoters of active genes have a nucleosome-free region (NFR) of about 150 bp overlapping the transcriptional start site and arrays of well-positioned and phased nucleosomes that radiate from the NFR (Fig. 4C). A notable reduction in apparent nucleosome occupancy extends up to 1 kb into the gene body. We also observed consistent nucleosome coordinates in an independent data set of H3K4me3-bearing nucleosomes[16] (Fig. 4D). Comparison of the nucleosome data (Fig 4D) with binding patterns of RNA Polymerase II[16] (Fig 4D) around active promoters indicates that phasing of positioned nucleosomes can be explained by packing of nucleosomes against Pol II stalled at the promoter, with Pol II potentially acting as the “barrier”. The set of inactive promoters, by contrast, exhibits neither a pronounced depletion of nucleosomes, nor a positioning and phasing signal (Fig. 4C). The transition of an inactive promoter to an active one is therefore likely to involve eviction of nucleosomes, coupled with positioning and phasing of nucleosomes neighboring RNA Pol II (Fig. 4E). These results suggest that CpG-rich segments in mammalian promoters override intrinsic signals of high nucleosome affinity (Supplementary Fig. 8) to become active; this would be in contrast to fly and yeast, where AT-rich promoters may comprise intrinsic sequence signals that are particularly prone to nucleosome eviction[28].

To explore how regulatory factors interact with sequence signals to influence nucleosome organization outside of promoters, we focused on binding sites of the NRSF repressor protein[15] and the insulator protein CTCF. NRSF and CTCF sites are flanked by arrays of

positioned nucleosomes (Fig. 4G, Supplementary Fig. 9), consistent with barrier-driven packing previously reported for CTCF[29,30]. Both proteins occupy additional linker space, with NRSF taking up an extra 37 bp and CTCF 74 bp. In agreement with sequence-based predictions[21], both CTCF and NRSF sites intrinsically encode high nucleosome occupancy as can be seen from the *in vitro* data (Fig. 4F, Supplementary Fig. 9), but this signal is overridden *in vivo* by occlusion of these sites from associating with nucleosomes. Additionally, phasing of nucleosomes around these regulatory sites is more compact in granulocytes compared to T-cells (Supplementary Fig. 9), again exemplifying the importance of cellular parameters for placement of nucleosomes.

Our genome-wide, deep sequence data of nucleosome positions facilitated an initial characterization of the determinants of nucleosome organization in primary human cells. Spacing of nucleosomes differs between cell types and between distinct epigenetic domains in the same cell type, and is influenced by transcriptional activity. We confirm positioning preferences in regulatory elements such as promoters and chromatin regulator binding sites, but find that the majority of the human genome exhibits little if any detectable positioning. The influence of sequence on positioning of nucleosomes *in vivo* is modest but detectable. Despite DNA sequence being a potent driver of nucleosome organization at certain sites, the cellular environment often overrides sequence signals and can drive nucleosomes to occupy intrinsically unfavorable DNA elements or evict nucleosomes from intrinsically favorable sites. We find evidence for the barrier model for nucleosome organization, and that barriers can be nucleosomes (positioned by container sites), RNA polymerase II (stalled at the promoter), or sequence-specific regulatory factors. Our nucleosome maps should be useful for investigating how nucleosome organization affects gene regulation and vice versa, as well as for pinpointing the mechanisms driving regional heterogeneity of nucleosome spacing.

Methods summary

Neutrophil granulocytes, CD4+ and CD8+ T-cells were isolated from donor blood using Histopaque density gradients and Ig-coupled beads against blood cell surface makers (pan T and CD4+ microbeads, Miltenyi Biotec). Nucleosome cores were prepared as previously described[7]; cells were snap-frozen and crushed to release chromatin, followed by micrococcal nuclease treatment. *In vitro* nucleosomes were prepared by combining human genomic DNA with recombinantly-derived histone octamers at an average ratio of 1 octamer per 850 bps. Unbound DNA was then digested using micrococcal nuclease. After digestion, reactions were stopped with EDTA, samples were treated with proteinase K, and nucleosome-bound DNA was extracted with phenol-chloroform and precipitated with ethanol (Supplementary methods). Purified DNA was size-selected (120–180 bp) on agarose to obtain mononucleosome cores, followed by sequencing library construction. RNA was isolated by homogenizing purified cells in Trizol, poly-A RNA was purified using Qiagen Oligotex kit and RNA-seq libraries were constructed using SOLiD Whole Transcriptome Analysis kit. All sequence data was obtained using SOLiD 35 bp protocol and aligned using the SOLiD pipeline against the human hg18 reference genome. Downstream analyses were all conducted using custom scripts (supplementary methods).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the Stanford Genetics/Pathology Sequencing Initiative. We would like to thank Geeta Narlikar for help with *in vitro* experiments, Life Technologies, especially Jason Briggs, for help with generating sequencing data, Phil Lacroute for help with sequence alignment, Stephen Galli for valuable discussions, Lia Gracey for critical reading of the manuscript, and members of Sidow and Fire labs for valuable feedback and discussions. Work in the Fire lab was partially supported by NIGMS (R01GM37706). AV was partially supported by an ENCODE subcontract to AS (NHGRI U01HG004695). SMJ was partially supported by the Stanford Genome Training program (NHGRI T32HG00044).

References and Notes

1. Mellor J. The dynamics of chromatin remodeling at promoters. Mol Cell. 2005; 19
2. Radman-Livaja M, Rando OJ. Nucleosome positioning: how is it established, and why does it matter? Dev Biol. 2010; 15
3. Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J, Segal E. The DNA-encoded nucleosome organization of a eukaryotic genome. Nature. 2009; 458
4. Berstein BE, Liu CL, Humphrey EL, Perlstein EO, Schreiber SL. Global nucleosome occupancy in yeast. Genome Biol. 2004; 5(9)
5. Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altshuler SJ, Rando O. Genome-scale identification of nucleosome positions in S. cerevisiae. Science. 2005; 22;309(5734)
6. Johnson SM, Tan FJ, McCullough HL, Riordan DP, Fire AZ. Flexibility and constraint in the nucleosome core landscape of Caenorhabditis elegans chromatin. Genome Research. 2006; 16(12)
7. Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, Zeng K, Malek JA, Costa G, McKernan K, Sidow A, Fire A, Johnson SM. A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Research. 2008; 18(7)
8. Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K. Dynamic regulation of nucleosome positioning in the human genome. Cell. 2010; 132(5)
9. Trifonov EN, Sussman JL. The pitch of chromatin DNA is reflected in it its nucleotide sequence. Proc Nat Acad Sci. 1980; 77
10. Kornberg RD. Sturcture of chromatin. Ann Rev Biochem. 1977; 46
11. Widom J. A relationship between the helical twist of DNA and the ordered positioning of nucleosomes in all eukaryotic cells. Proc Natl Acad Sci U S A. 1992; 1;89(3)
12. Schlegel RA, Haye KR, Litwack AH, Phelps BM. Nucleosome repeat lengths in the definitive erythroid series of the adult chicken. Biochim Biphys Acta. 1980; 606
13. Fan Y, Nikitina T, Zhao J, Fleury TJ, Bhattacharyya R, Bouhassira EE, Stein A, Woodcock CL, Skoultchi AI. Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. Cell. 2005; 123
14. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008; 5
15. Valouev A, Johnson DS, Sundquist A, Medina C, Anton E, Batzoglou S, Myers RM, Sidow A. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nat Methods. 2009; 5
16. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. High-resolution profiling of histone methylations in the human genome. Cell. 2007; 129
17. Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, Zhao K. Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet. 2008; 40

18. Hughes A, Rando O. Chromatin programming by sequence is there more to the nucleosome code than %GC? J Biol. 2009; 8(11)
19. Satchwell SC, Drew HR, Travers AA. Sequence periodicities in chicken nucleosome core DNA. J Mol Biol. 1986; 20;191(4)
20. Segal E, Fondufe-Mittendorf Y, Chen L, Thastrom A, Field Y, Moore IK, Wang JP, Widom J. A genomic code for nucleosome positioning. Nature. 2006; 442(7104)
21. Tillo D, Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Field Y, Lieb JD, Widom J, Segal E, Hughes TR. High nucleosome occupancy is encoded at human regulatory sequences. PLoS One. 2010; 5
22. Mavrich TN, Ioshikhes IP, Venters BJ, Jiang C, Tomsho LP, Qi J, Schuster SC, Albert I, Pugh BF. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Gnome Res. 2008; 18
23. Mavrich TN, Jiang C, Ioshikhes IP, Li X, Venters BJ, Zanton SJ, Tomsho LP, Qi J, Glaser RL, Schuster SC, Gilmour DS, Albert I, Pugh BF. Nucleosome organization in the Drosophila genome. Nature. 2008; 453
24. Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C. A high-resolution atlas of nucleosome occupancy in yeast. Nat Genet. 2007; 39(10)
25. Gu SGA. Fire Partitioning the C. elegans genome by nucleosome modification, occupancy, and positioning. Chromosoma. 2010; 119(1)
26. Sasaki S, Mello CC, Shimada A, Nakatani Y, Hashimoto S, Ogawa M, Matsushima K, Gu SG, Kasahara M, Ahsan B, Sasaki A, Saito T, Suzuki Y, Sugano S, Kohara Y, Takeda H, Fire A, Morishita S. Chromatin-associated periodicity in genetic variation downstream of transcriptional start sites. Science. 2009; 323
27. Zhang Y, Moqtaderi Z, Rattner BP, Euskirchen G, Snyder M, Kadonaga JT, Liu XS, Struhl K. Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. Nat Struct Mol boil. 2009; 16(8)
28. Field Y, Fondufe-Mittendorf Y, Moore IK, Mieczkowski P, Kaplan N, Lubling Y, Lieb JD, Widom J, Segal E. Gene expression divergence in yeast is coupled to evolution of DNA-encoded nucleosome organization. Nat Genet. 2009; 41(4)
29. Chuddapah S, Jothi R, Schones DE, Roh TY, Cui K, Zhao K. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. Genome Res. 2009; 19
30. Fu Y, Sinha M, Peterson CL, Weng Z. The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. PLoS Genet. 2008; 4

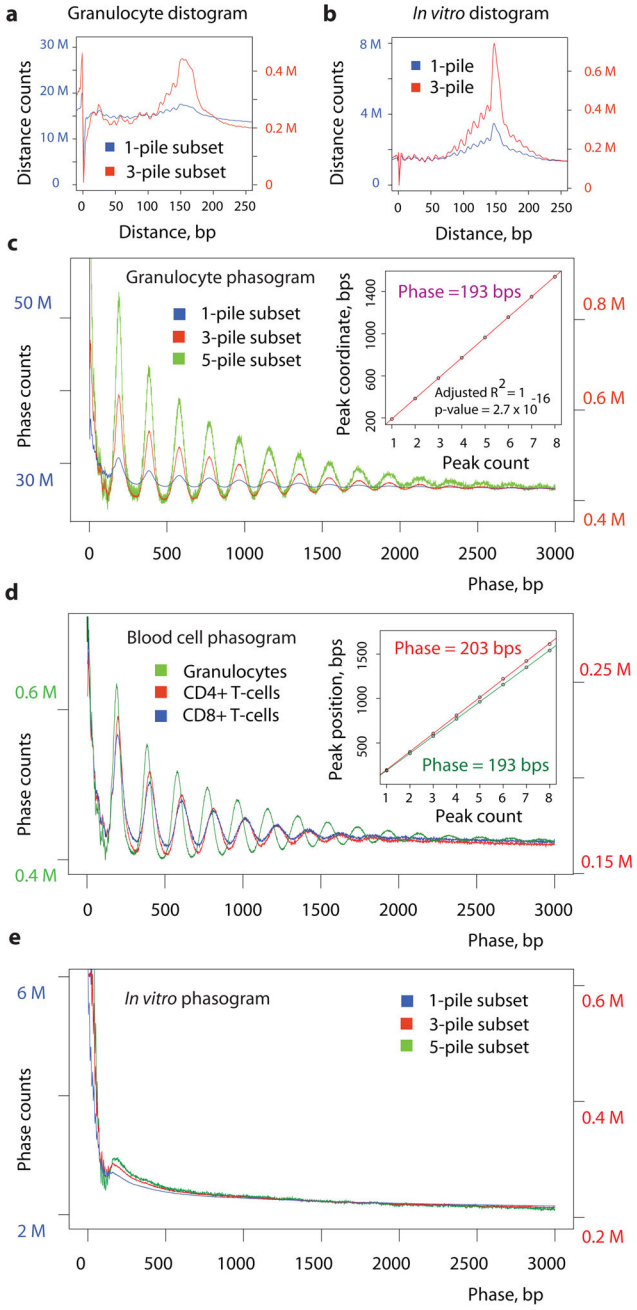

Figure 1. Global parameters of cell-specific nucleosome phasing and positioning in human

a, *In vivo* granulocyte distogram (calculation explained in Supplementary Fig. 3A). X axis represents the range of recorded distances. Y axis represents frequencies of observed distances within 1-pile (blue) and 3-pile (red) subsets. 1-pile subset represents the entire dataset, 3-pile subset represents a subset of sites containing 3 or more coincident read starts. **b,** Distogram of the *in vitro* reconstituted nucleosomes showing 1-and 3-pile subsets as in (a). **c,** *In vivo* granulocyte phasogram (calculation explained in Supplementary Fig. 3B). X axis shows the range of recorded phases. Y axis shows frequencies of corresponding phases. Plotted are phasograms of 1-, 3-, and 5-pile subsets. Inset, linear fit to the positions of the

phase peaks within 3-pile subsets (slope = 193 bp). **d,** Phasograms of blood cell types. Inset, linear fits in CD4+ T-cells (203 bp) and granulocytes (193 bp). **e,** Phasograms of 1-, 3-and 5-pile subsets in the *in vitro* data.

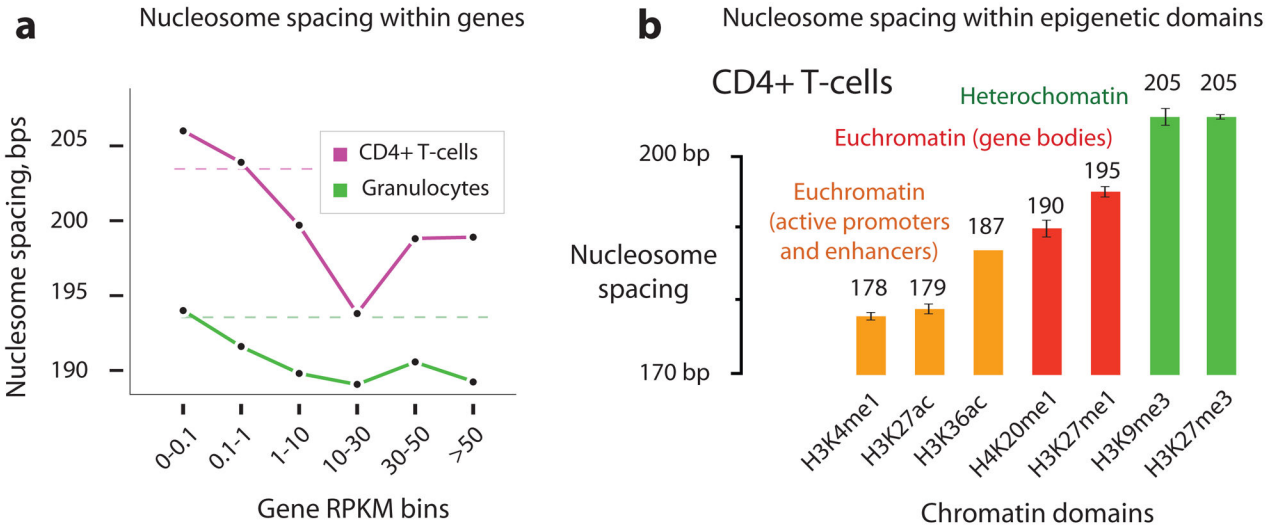

Figure 2. Transcription and chromatin modification-dependent nucleosome spacing

a, Nucleosome spacing as a function of transcriptional activity. X axis represents gene expression values binned according to RPKM values. Internucleosome spacing is plotted along the Y axis. Dashed lines represent genome-wide average spacing for each cell type. **b,** Nucleosome spacing within genomic regions marked by specific histone marks in CD4+ T-cells. Bar height plots estimated nucleosome spacing for each histone modification. Bar colors differentiate chromatin types (euchromatin vs heterochromatin).

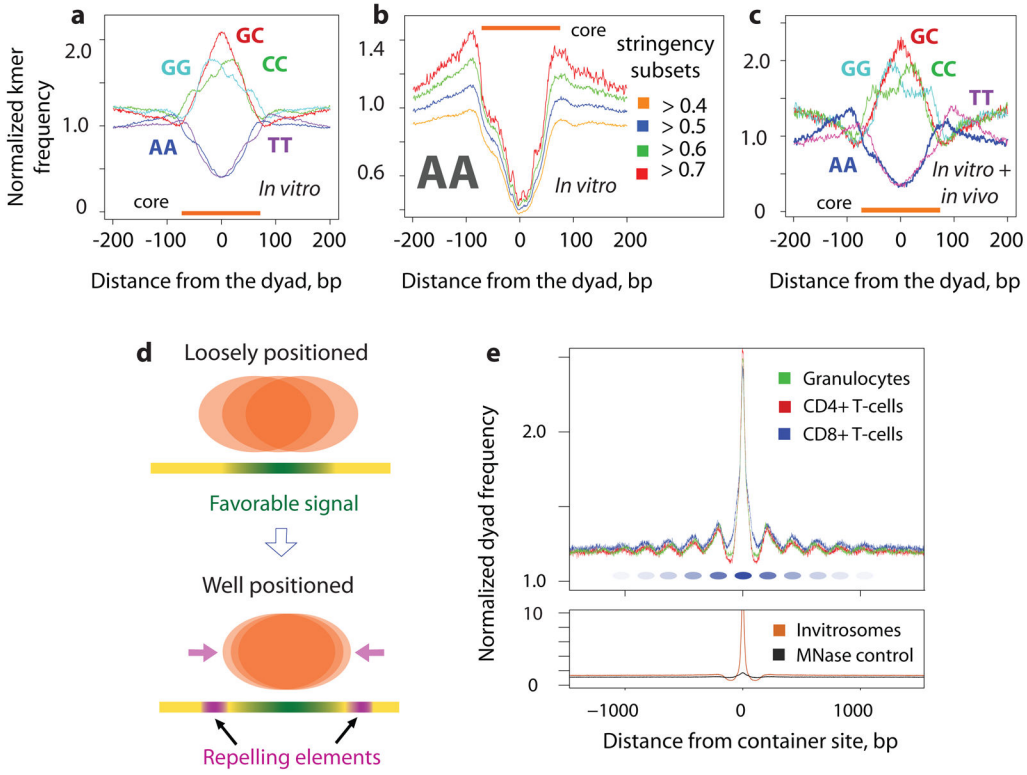

Figure 3. Sequence signals that drive nucleosome positioning

a, Sequence signals within sites containing moderately positioned *in vitro* nucleosomes (stringency > 0.5). Distance from the positioned dyad to a given dinucleotide are plotted along the X axis; Y axis represents frequency of a given k-mer divided by its genome-wide expectation. The 147 bp footprint of a nucleosome is indicated by an orange band. **b,** Changes in AA dinucleotide usage with increasing positioning stringency. X and Y axes same as in (a). Shown are curves of AA usage within the sites of increasingly positioned dyads (stringency cutoffs of 0.4, 0.5, 0.6, 0.7). **c,** Sequence signals within sites containing *in vitro* positioned nucleosomes (stringency > 0.5) that also have high *in vivo* stringency (stringency > 0.4). X and Y axes same as in (a). **d,** Schematic depiction of the container site positioning mechanism. The C/G-rich core area (green) favors occupancy, but does not precisely position the nucleosome (top). Adding flanking A/T-rich repelling elements (purple, bottom) restricts the position of the nucleosome. **e,** Nucleosome organization around container sites *in vivo* and *in vitro*. X axis represents distances from the dyads to container sites(based on 300,000 container sites). Frequencies of nucleosome dyads around those sites are plotted along the Y axis. The upper plot shows distribution of *in vivo* dyads across CD4+ cells, CD8+ cells, and granulocytes. The ovals depict hypothetical nucleosome positions across the site with color intensities reflecting their positioning strength. The lower plot shows distribution of dyads *in vitro* and in MNase control.

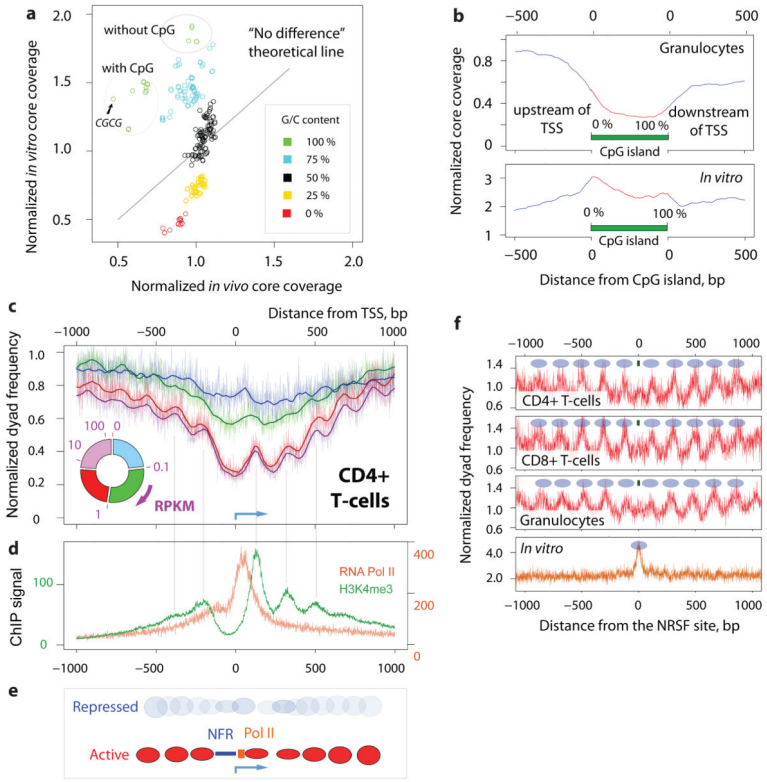

Figure 4. Influence of gene regulatory function on nucleosome positioning

a, Comparison of sequence preferences of nucleosomes *in vivo* and *in vitro*. Normalized nucleosome core coverage *in vivo* (granulocytes) for a given sequence 4-mer are plotted along the X axis. *In vitro* core coverage is plotted along the Y axis. Each data point on the plot represents one of the 256 possible 4-mers (colored according to their G/C content). The diagonal line depicts the positions in the plot for which sequence-based preferences of nucleosomes would be the same *in vivo* and *in vitro*. **b,** Nucleosome core coverage over CpG islands *in vivo* and *in vitro*. X axis represents coordinates within CpG islands (0–100%) and flanking upstream of the transcriptional start sites (TSS) (left) and downstream of the TSS (right). Normalized frequencies of nucleosome cores *in vivo* (upper plot) and *in vitro* (lower plot)are plotted along the Y axis. **c,** *In vivo* CD4+ T-cell nucleosome organization around promoters. X axis represents distance from the TSS (blue arrow). Normalized frequencies of nucleosome dyads are plotted along the Y axis. Shown are nucleosome arrangements within four gene groups (not expressed 0–0.1 RPKM, low expressed 0.1–1 RPKM, moderately expressed 1–8 RPKM, highly expressed >8 RPKM). Pie chart depicts distribution of RPKM values across gene groups. **d,** RNA Pol II binding signal within highly expressed genes (orange curve) and H3K4me3-marked nucleosome dyad frequency (green curve) within highly expressed genes(>8 RPKM). Nucleosomes show consistent positions, indicated by grey lines pointing to nucleosome centers. **e,** Schematic depiction of nucleosome organization around promoters of repressed and active genes. Promoters of repressed genes do not have a well-defined nucleosome organization, while promoters of active genes have a nucleosome free region (NFR, blue), RNA Pol II (orange) localized at the NFR boundary, and positioned nucleosomes (red) radiating from the NFR. Height of the

ovals represents nucleosome frequency(inferred from c). **f,** Nucleosome distribution around the top 1000 NRSF sites *in vivo* and *in vitro*. Distances from the NRSF binding sites are plotted along the X axis. Y axis represents the normalized frequency of nucleosome dyads. Blue ovals depict hypothetical nucleosome positions. NRSF binding site is shown by the green rectangle.