LETTER

Corrected: Publisher Correction

https://doi.org/10.1038/s41586-018-0567-3

Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing

Binbin Lai[1], Weiwu Gao[1,2], Kairong Cui[1], Wanli Xie[1,3], Qingsong Tang[1], Wenfei Jin[4], Gangqing Hu[1], Bing Ni[2] & Keji Zhao[1]*

Nucleosome positioning is critical to chromatin accessibility and is associated with gene expression programs in cells[1–3]. Previous nucleosome mapping methods assemble profiles from cell populations and reveal a cell-averaged pattern: nucleosomes are positioned and form a phased array that surrounds the transcription start sites of active genes[3–6] and DNase I hypersensitive sites[7]. However, even in a homogenous population of cells, cells exhibit heterogeneity in expression in response to active signalling[8,9] that may be related to heterogeneity in chromatin accessibility[10–12]. Here we report a technique, termed single-cell micrococcal nuclease sequencing (scMNase-seq), that can be used to simultaneously measure genome-wide nucleosome positioning and chromatin accessibility in single cells. Application of scMNase-seq to NIH3T3 cells, mouse primary naive CD4 T cells and mouse embryonic stem cells reveals two principles of nucleosome organization: first, nucleosomes in heterochromatin regions, or that surround the transcription start sites of silent genes, show large variation in positioning across different cells but are highly uniformly spaced along the nucleosome array; and second, nucleosomes that surround the transcription start sites of active genes and DNase I hypersensitive sites show little variation in positioning across different cells but are relatively heterogeneously spaced along the nucleosome array. We found a bimodal distribution of nucleosome spacing at DNase I hypersensitive sites, which corresponds to inaccessible and accessible states and is associated with nucleosome variation and variation in accessibility across cells. Nucleosome variation is smaller within single cells than across cells, and smaller within the same cell type than across cell types. A large fraction of naive CD4 T cells and mouse embryonic stem cells shows depleted nucleosome occupancy at the de novo enhancers detected in their respective differentiated lineages, revealing the existence of cells primed for differentiation to specific lineages in undifferentiated cell populations.

To understand the principles that underline chromatin heterogeneity as related to nucleosome positioning and chromatin accessibility, we developed the scMNase-seq technique to simultaneously measure nucleosome positioning and chromatin accessibility in single cells. We applied scMNase-seq to 48 NIH3T3 single cells, 198 mouse embryonic stem cells (ESCs), and 278 mouse naive CD4 T cells, obtaining on average about 3, 0.9 and 0.7 million unique fragments, respectively, for each cell type (Fig. 1a, Supplementary Table 1). Sequence reads from sorted human or mouse cells from a mixed population mapped exclusively to the respective genome, suggesting that there was no DNA contamination across cells (Extended Data Fig. 1a). Pooled single-cell reads revealed a size distribution that was consistent with that obtained by bulk-cell MNase-seq (Extended Data Fig. 1b). We considered fragments with a length between 140 and 180 bp as canonical nucleosomes, and fragments with a length $\leq$ 80 bp as subnucleosome-sized particles (Extended Data Fig. 1b, c). Compared to CD4 T cells and mouse ESCs (Fig. 1b), NIH3T3 libraries have the largest number of non-redundant reads (Extended Data Fig. 1d) and the highest genomic coverage (5–30%) of nucleosomes—probably owing to the polyploidy of NIH3T3 cells (Extended Data Fig. 1e). Nevertheless, all three cell types have a similar nucleosome density across different genomic regions, which suggests that representation of the genome is relatively even (Extended Data Fig. 1f). The nucleosome positioning and the enrichment of subnucleosome-sized particles surrounding DNase I hypersensitive sites (DHSs), the transcription start sites (TSSs) of active genes and CTCF-binding sites were consistent between pooled scMNase-seq and bulk-cell MNase-seq data (Fig. 1c, Extended Data Fig. 2a–h). The density of subnucleosome-sized particles from pooled single cells is correlated with the DNase I tag density at DHSs and with gene expression at TSSs, suggesting that subnucleosome-sized particles are predictive of chromatin accessibility (Extended Data Fig. 2i, j). Moreover, the percentage of DHSs detected by scMNase-seq was higher than that detected by single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq)[10] with the same sequencing redundancy (owing to the higher complexity and non-redundant read-number of scMNase-seq libraries), although when using scMNase-seq the percentage of recovered DHSs per non-redundant read for subnucleosome-sized particles was relatively lower than scATAC-seq fragments (Extended Data Fig. 2k–n). Nucleosome positions from single cells, aggregated nucleosome density from pooled single cells and tag density from bulk-cell MNase-seq at representative cell-type-specific genes are shown for all three cell types (Fig. 1d). Notably, the similarity of aggregated nucleosome profiles between pooled single cells and bulk cells is correlated with nucleosome positioning stringency and nucleosome coverage, and is higher for active promoters than it is for silent promoters (Fig. 1d, Extended Data Fig. 2o). These results demonstrate that scMNase-seq can simultaneously measure nucleosome positioning and chromatin accessibility in single cells.

Although nucleosome positioning[13] is well-studied[5,14–16] on the basis of large numbers of pooled cells, genome-wide nucleosome spacing patterns are poorly understood because current knowledge about nucleosome spacing is limited to the positioned nucleosomes[7,17,18]. We profiled the distribution of nucleosome-to-nucleosome distance for different single cells and used relative peak height to measure the uniformity of nucleosome spacing for both positioned and non-positioned nucleosomes (Extended Data Fig. 3a–c, Supplementary Methods). This analysis revealed a high degree in spacing uniformity in single cells regardless of positioning stringency; decreased uniformity in spacing was observed as positioning stringency decreased, when using either the pooled single cells or bulk-cell MNase-seq data[19] (Extended Data Fig. 3d). The bulk-cell MNase-seq data failed to reveal the actual spacing pattern owing to the mixture of non-positioned nucleosomes from a population of different cells.

The degree of uniformity in spacing in the promoter regions of silent genes is higher than that of active genes (Fig. 2a, b, Extended Data Fig. 4a, b), and uniformity is higher in non-DHS than in DHS regions (Fig. 2c, d, Extended Data Fig. 4a–c). Notably, the higher uniformity of spacing in non-DHS regions was also observed in single haploid mouse ESCs and haploid chromosome X in single mouse ESCs (Extended Data

[1]Laboratory of Epigenome Biology, Systems Biology Center, National Heart, Lung and Blood Institute, NIH, Bethesda, MD, USA. [2]Department of Pathophysiology and High Altitude Pathology, Third Military Medical University, Chongqing, China. [3]Institute of Anesthesiology and Critical Care Medicine, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. [4]Department of Biology, South University of Science and Technology of China, Shenzhen, China. *e-mail: zhaok@nhlbi.nih.gov

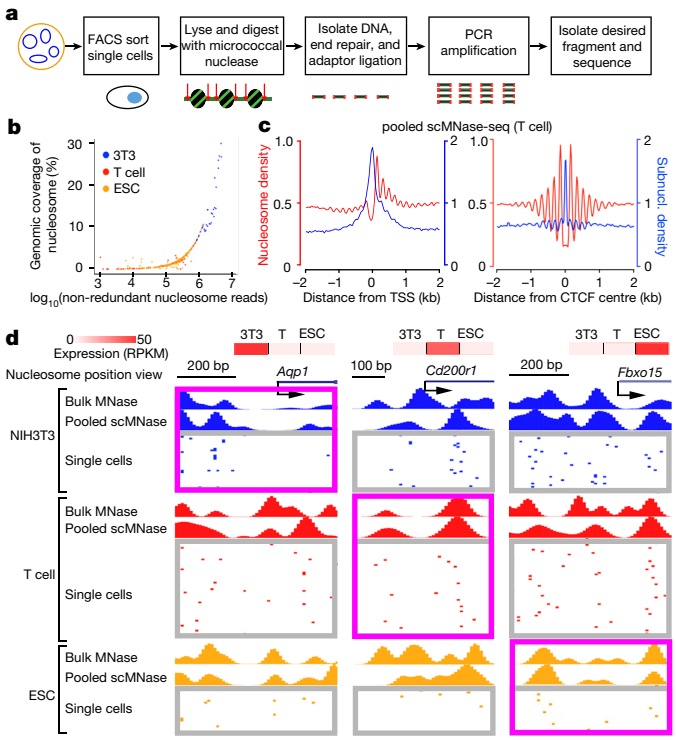


Fig. 1 | scMNase-seq simultaneously measures the positions of nucleosomes and subnucleosome-sized particles in single cells. **a**, Schema of scMNase-seq. **b**, Plot of non-redundant nucleosome read number (*x* axis) and genomic coverage of nucleosomes (*y* axis) for single NIH3T3 cells, CD4 T cells and mouse embryonic stem cells. **c**, Average density profiles of nucleosomes (red) and subnucleosome-sized particles (blue) relative to TSS of active genes (left) and CTCF-binding sites (right) for pooled CD4 T cells scMNase-seq data. Subnucl., subnucleosome-sized particles. **d**, Genome browser view of single-cell nucleosome positions for NIH3T3 cells, CD4 T cells and mouse ESCs at TSSs of three representative cell-type-specific gene loci. Single-cell libraries that have at least one nucleosome within any of three genomic regions are shown. Tracks for tag density of corresponding bulk-cell MNase-seq data (one representative from two repeated experiments is shown) and pooled scMNase-seq data (all single cell libraries with detected nucleosomes in selected genomic regions are included) are also shown. The nucleosome maps at expressed genes for each cell type are highlighted with pink rectangle. The expression levels of genes are shown in the heat map above the tracks. 3T3, NIH3T3 cells; T, CD4 T cells; ESC, mouse ESC; RPKM, reads per kilobase of transcript per million mapped reads.

Fig. 4d–g), and was independent of MNase concentration (Extended Data Fig. 4h–m). Furthermore, nucleosome spacing in active chromatin regions associated with H3K4me1, H3K4me3, H3K27ac, H3K9ac and H2AZ shows a lower degree of uniformity than transcribed regions marked by H3K36me3, heterochromatic regions marked by H3K27me3 or not marked by any of the histone modifications that we studied (Extended Data Fig. 4n, o).

We next measured variation in nucleosome positioning not only across cells but also within single cells (across different alleles) by calculating the mean value of distances between two overlapping nucleosomes within genomic regions related to a particular feature (for example, active promoters) (Extended Data Fig. 5a). As expected, variation in nucleosome positioning around the TSSs of active genes—where nucleosomes are phased relative to TSS—is smaller than that around the TSSs of silent genes (Fig. 2e). In addition, nucleosome positions show smaller variation at the centre of DHSs and the centre of chromatin regions enriched in active histone modifications than they do elsewhere (Extended Data Fig. 5b–g).

The results above reveal that there are different rules of nucleosome organization in different chromatin regions. In silent chromatin states—such as in repressed promoters and heterochromatic regions—nucleosomes are highly uniformly spaced, but are not positioned relative to the underlying genomic DNA across different arrays. By contrast, in active chromatin states—such as transcribed promoters and DHS regions—nucleosomes are positioned but are not as uniformly spaced (Fig. 2f). This model was further supported by the observation that nucleosomes in promoter regions of silent genes, non-DHS regions and heterochromatic regions show higher synchronized shift scores than nucleosomes in promoters of active genes, DHS regions and regions marked by active histone modifications (Extended Data Fig. 6a, b). Furthermore, the synchronized shift score is dependent on nucleosome spacing; the highest scores are in the spacing range of 180–185 bp, which is dominant throughout the genome in all single cells (Extended Data Fig. 6c, d). The nucleosome spacing might indicate a stable structure for packaging nucleosomes[20] in silent chromatin states, which is probably collectively determined by chromatin assembly factors[21], linker histones[22,23] and the environment surrounding chromatin fibres. In active states, ATP-dependent chromatin remodelling activities[15,24] may reposition nucleosomes[1,25] and consequently change the local nucleosome spacing to facilitate chromatin accessibility and gene transcription. Notably, the average nucleosome spacing surrounding the DHSs is shorter than that in non-DHS regions (Extended Data Fig. 6e, f), which may be the result of repositioning of the nucleosomes to allow accessibility of the DHS regions.

Although nucleosomes are positioned surrounding DHSs to ensure chromatin accessibility[7], extensive heterogeneity of chromatin accessibility across different single cells[10,12] implies heterogeneity of nucleosome positioning at the same DHS. Profiling nucleosome-to-nucleosome distances over DHSs reveals two distinct peak patterns: one has a summit at about 190 bp and the other has a summit at about 300 bp, which presumably corresponds to two different chromatin states (closed or open) (Fig. 3a, b). More than 80% of the DHSs have both spacing types at the same DHS in different single cells (Fig. 3c), and the higher DNase I tag density at DHSs measured in bulk cells[26] is associated with more wide-spacing DHSs in single cells (Extended Data Fig. 7a). Furthermore, the DHSs with a higher fraction of wide space—which is not related to MNase digestion—are associated with higher accessibility, when measured by bulk-cell DNase I hypersensitive sites sequencing (DNase-seq) or by scMNase-seq subnucleosome-sized particles (Extended Data Fig. 7b–d), and with lower variation in DHS accessibility and nucleosome positioning across different single cells (Fig. 3d, e). These results indicate that one DHS may have two types of nucleosome organization (wide or narrow spacing) across different single cells; the degree of accessibility of a DHS as well as the variation in DHS accessibility and nucleosome positioning across cells are directly linked to the ratio between the two states of nucleosome organization in different single cells.

Furthermore, variation in nucleosome positioning around DHSs is positively correlated with variation in accessibility across different single cells (Fig. 3f). The fraction of single cells with nucleosomes positioned around DHSs is correlated with the number of cells detected as DHSs (Extended Data Fig. 7e). The variation in nucleosome positioning around TSSs in different single cells is also correlated with variation in gene expression. The TSSs with +1 nucleosomes that show higher variation in nucleosome positioning also show higher variation in expression across different single cells (Fig. 3g). Genes for which expression was detected in a higher fraction of single cells exhibit positioned +1 nucleosomes in a higher fraction of single cells than do the genes with a lower fraction of expression (Extended Data Fig. 7f). The top 1,000 active genes with smallest nucleosome variance around their TSS across cells are enriched in common biological processes such as translation and protein transport (Extended Data Fig. 7g), consistent with the notion that house-keeping genes display less variation in nucleosome positioning. Furthermore, variation within a cell in nucleosome positioning around DHSs, or at the +1 nucleosome of the TSSs of active genes, is smaller than that across different single cells (Extended Data Fig. 7h, i). The variation in nucleosome positioning within the cell type is smaller than that across different cell types

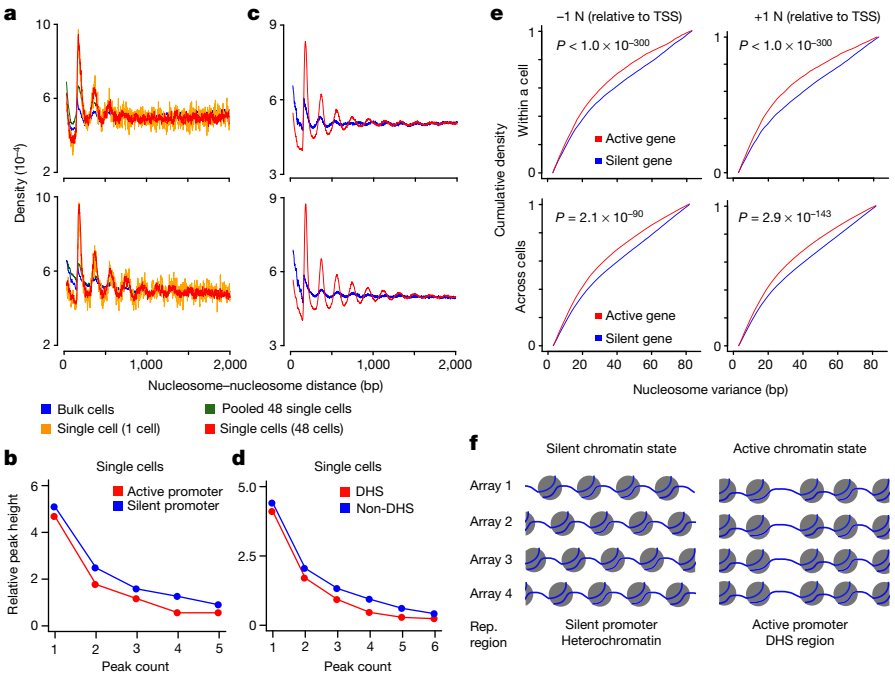


Fig. 2 | Profiling nucleosome positioning and spacing in single cells reveals distinct nucleosome organization principles at active and silent chromatin regions. a, Density plots of nucleosome-to-nucleosome distance within active-gene promoters (top) and silent-gene promoters (bottom) for bulk-cell MNase-seq, pooled 48 NIH3T3 single cells, one representative single cell and 48 single-cell scMNase-seq datasets. **b**, The relative peak heights based on the data from **a** reveal a higher degree of uniformity in spacing within silent-gene promoters than active-gene promoters. **c**, Density plots of nucleosome-to-nucleosome distance within DHS regions (top) and non-DHS regions (bottom) for bulk-cell MNase-seq (blue) and 48 single-cell scMNase-seq (red) datasets. **d**, The relative peak heights based on the data from **c** reveal a higher degree of uniformity in spacing within non-DHS regions than DHS regions. **e**, Cumulative density of variance in nucleosome positioning in active and silent genes within a cell (top) and across single cells (bottom), at −1 (left) and +1 (right) nucleosomes relative to the TSS. Top left, $n = 7{,}574$ and 13,107 nucleosome pairs for active and silent genes, respectively; bottom left, $n = 164{,}512$ and 304,847 nucleosome pairs for active and silent genes, respectively; top right, $n = 11{,}388$ and 17,631 nucleosome pairs for active and silent genes, respectively; bottom right, $n = 237{,}006$ and 416,328 nucleosome pairs for active and silent genes, respectively. *P* values were calculated using one-sided Mann–Whitney *U*-test. **f**, Cartoon illustrating nucleosome organization patterns in silent (left) and active (right) chromatin states. Rep., representative genomic region.

(Extended Data Fig. 7j). Clustering based on similarity in nucleosome positioning at all the DHSs across all the single cells from three cell types separated these single cells into three clusters that correspond to the respective cell types—this clustering is independent of experiment time and fragment-size ratio (Extended Data Fig. 7k).

The DNA sequence has an important role in nucleosome positioning[4,14,16]. Consistent with a previous report[14], we observed high CC, GG and GC frequency in nucleosome-occupied sequences and high AA, TT, AT and TA frequency in flanking regions in single cells, as well as a periodical pattern that supports the rotational positioning of nucleosomes[4,16] (Extended Data Fig. 8a). Smaller variation in nucleosome positioning is associated with lower frequencies of CC, GG and GC and higher frequencies of AA, TT, AT and TA in the flanking region (Extended Data Fig. 8b–e). We next explored the relationship between variance in DNA sequence and variance in nucleosome positioning. Our analysis shows that sequences occupied by nucleosomes have a higher fraction of alternative bases than those that are occupied by subnucleosome-sized particles, by tags from DNase-seq or by tags from CTCF chromatin immunoprecipitation with sequencing (ChIP-seq) (Extended Data Fig. 8f, g), which supports the notion that sequence variants influence transcription-factor binding[27] and nucleosome positioning[28]. We found that single-base variance within nucleosome regions is positively correlated with nucleosome variance across cells (Extended Data Fig. 9h). Furthermore, the single-base variance at transcription-factor motifs is positively correlated with nucleosome variance at DHSs and is also positively correlated with gene expression variation across different single cells (Extended Data Fig. 9i, j).

Enhancers display remarkable cell-type specificity. Consistent with a previous observation that active enhancers are associated with a nucleosome loss[3], the naive CD4 T cell-specific enhancers displayed decreased nucleosome occupancy in naive CD4 T cells, as revealed by the pooled scMNase-seq data from naive CD4 T cells; by contrast, enhancers that are specific to T helper 1 (T_H1) and T helper 2 (T_H2) cells showed only a very minor overall nucleosome loss in naive CD4 T cells (Extended Data Fig. 9a, b). However, examination of the nucleosome patterns at the T_H1- and T_H2-specific enhancers across different single cells revealed that 19% and 29% of naive CD4 T cells showed decreased nucleosome occupancy—which is independent of fragment-size ratio—at the de novo enhancers of T_H1 and T_H2 cells, respectively, whereas much smaller fractions of mouse ESCs and NIH3T3 cells showed decreased nucleosome occupancy at these enhancers (Fig. 4a, Extended Data Fig. 9c–e). Furthermore, subgroups of T cells that show decreased nucleosome occupancy at the T_H1 and T_H2 enhancers do not have much overlap (Extended Data Fig. 9f), which suggests they are specifically primed for the corresponding lineages. The T_H1-specific enhancers with the most nucleosome loss in naive CD4 T cells are linked to genes that encode T_H1 cytokine (*Ifng*) and key regulators (*Tbx21*, *Stat1* and *Stat4*) (Extended Data Fig. 9g, h); the T_H2-specific enhancers with the most nucleosome loss are linked to genes that encode key regulators for T_H2 differentiation (*Il4* and *Stat6*) (Extended Data Fig. 9i, j). Motif analysis revealed that the nucleosome loss at T_H1 enhancers is specifically associated with motifs for RELA, which promotes T_H1 differentiation; the nucleosome loss at T_H2 enhancers is specifically associated with motifs for GATA3 and STAT6, which promote T_H2 differentiation (Extended Data Fig. 9k). Gene Ontology analysis revealed that the higher-ranked nucleosome losses at both T_H1 and T_H2 enhancers are associated with functions in T cell differentiation, immune system process and cytokine production (Extended Data Fig. 9l, m). These results suggest that a large fraction of naive CD4 T cells have already experienced differentiating

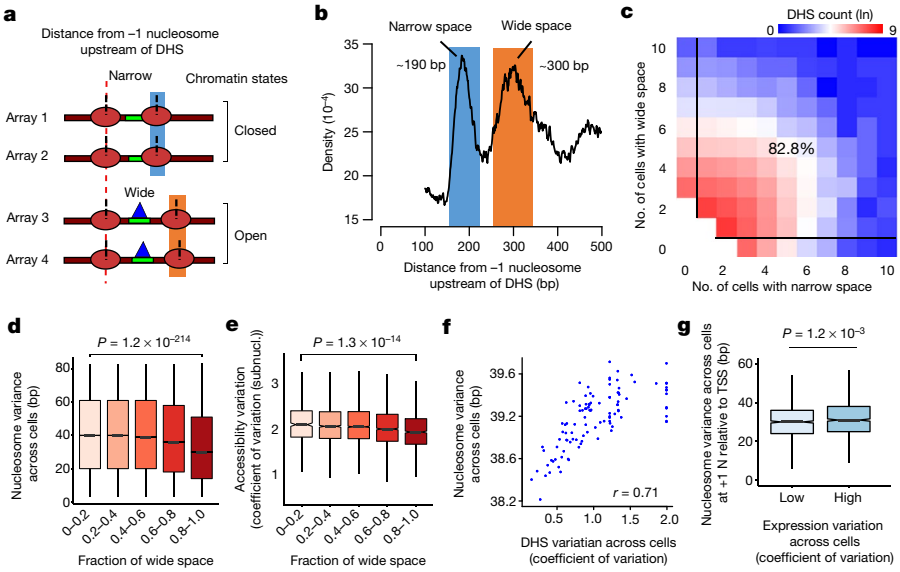


Fig. 3 | The bimodal distribution of nucleosome spacing across DHSs is associated with the cell-to-cell variation in nucleosome positioning and chromatin accessibility. a, Schema of nucleosome spacing across a DHS and two chromatin states inferred by nucleosome spacing. **b**, Density plot of nucleosome spacing across a DHS within single cells reveals two peaks that correspond to narrow spacing (blue) and wide spacing (red). **c**, Heat map showing DHS frequency as a function of number of cells with narrow spacing and number of cells with wide spacing. The percentage of DHSs in which there are both types of spacing across a DHS in different single cells is shown. **d**, **e**, Box plots showing the cell-to-cell variation in nucleosome positioning (**d**) and chromatin accessibility (**e**) for five groups of DHSs, defined by fraction of wide space. Data represent 612, 2,088, 3,858, 2,500 and 1,586 DHSs (from left to right). **f**, Scatter plot showing nucleosome variance (*y* axis) and DHS variation (*x* axis) across cells for 106 bins of DHSs, based on DHS variation. Each dot represents the average of 500 DHSs for each bin. Pearson's correlation was calculated. **g**, Box plot showing nucleosome variation at +1 nucleosome relative to TSS for two groups of genes sorted by expression variation. Low, bottom 25% ($n = 1{,}171$ genes); high, top 25% ($n = 1{,}174$ genes). In **d**, **e** and **g**, *P* values were calculated by one-sided Mann–Whitney *U*-test. In the box plots, centre line is median; boxes, first and third quartiles; whiskers, 1.5× interquartile range; notch, 95% confidence interval of the median.

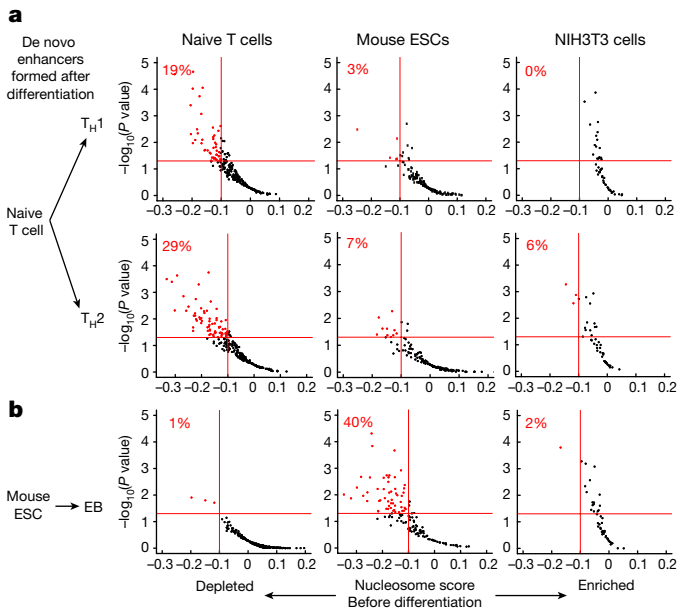


Fig. 4 | A subgroup of undifferentiated cells shows a nucleosome signature primed for differentiation. a, **b**, A large fraction of naive CD4 T cells shows decreased nucleosome occupancy at the de novo enhancers that are formed either in T_H1 (**a**, top) or T_H2 cells (**a**, bottom), whereas only a small fraction of mouse ESCs and NIH3T3 cells shows nucleosome depletion at the same enhancers. By contrast, a large fraction of mouse ESCs shows depleted nucleosomes at the de novo enhancers that are formed in EBs, whereas only a small fraction of naive CD4 T cells and NIH3T3 cells shows nucleosome depletion at the same enhancers. The fractions of primed cells are shown in red. Data represent 237 single naive T cells, 143 single mouse ESCs and 48 single NIH3T3 cells.

signalling events during the developmental history of these cells, which have primed the de novo enhancers of T_H1 or T_H2 cells by means of decreased nucleosome occupancy in the undifferentiated naive CD4 T cells.

Similarly, mouse ESCs displayed a substantial nucleosome loss at the mouse ESC-specific enhancers but only a minor loss at embryoid-body (EB)-specific enhancers, which are formed de novo after differentiation from mouse ESCs (Extended Data Fig. 10a, b). Analysis of single cells revealed that 40% of mouse ESCs showed decreased nucleosome occupancy at the de novo EB-specific enhancers, whereas only 1% and 2% of naive CD4 T cells and NIH3T3 cells, respectively, showed decreased nucleosome occupancy at these enhancers (Fig. 4b, Extended Data Fig. 10c, d). The EB enhancers with the most nucleosome loss are linked to genes that include mesoderm markers (*Brachyury* (also known as *T*) and *Wnt3*) and endoderm markers (*Gata4* and *Gata6*) (Extended Data Fig. 10e, f), and are associated with stem cell differentiation and development of various lineages, such as myeloid, neural tube and placental cells (Extended Data Fig. 10g). These results reveal the heterogeneity of cultured mouse ESCs, and suggest that some of these cells are already primed for differentiation by the reorganization of their nucleosome structure at enhancers formed in the differentiating EBs.

Here we introduce scMNase-seq, a powerful method for simultaneously measuring chromatin accessibility and nucleosome positioning in single cells, which may be paired with existing approaches—such as single-cell RNA-seq[9], single-cell DNase-seq[12] and/or single-cell ChIP-seq[29]—for systems analysis and to provide further insights into the molecular underpinning of cellular heterogeneity. Our application of scMNase-seq to three types of single cells revealed principles of nucleosome organization in different chromatin regions as well as heterogeneity of nucleosome positioning and spacing at DHSs. Our data suggest that the cellular heterogeneity of undifferentiated

cells is related to heterogeneous nucleosome organization in critical regulatory regions, which reflects the differentiation potential of these cells.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Code availability

Custom codes for the quantification of the uniformity of nucleosome spacing and calculation of nucleosome occupancy score are available at https://github.com/binbinlai2012/scMNase.

Data availability

The scMNase-seq datasets have been deposited in the Gene Expression Omnibus database with accession number GSE96688.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0567-3.

Received: 20 March 2017; Accepted: 20 August 2018;
Published online 26 September 2018.

1. Shivaswamy, S. et al. Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biol.* **6**, e65 (2008).
2. Lam, F. H., Steger, D. J. & O'Shea, E. K. Chromatin decouples promoter threshold from dynamic range. *Nature* **453**, 246–250 (2008).
3. Schones, D. E. et al. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887–898 (2008).
4. Albert, I. et al. Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature* **446**, 572–576 (2007).
5. Mavrich, T. N. et al. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res.* **18**, 1073–1083 (2008).
6. Mavrich, T. N. et al. Nucleosome organization in the *Drosophila* genome. *Nature* **453**, 358–362 (2008).
7. Gaffney, D. J. et al. Controls of nucleosome positioning in the human genome. *PLoS Genet.* **8**, e1003036 (2012).
8. Tang, F. et al. mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
9. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
10. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
11. Cusanovich, D. A. et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
12. Jin, W. et al. Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. *Nature* **528**, 142–146 (2015).
13. Kaplan, N., Hughes, T. R., Lieb, J. D., Widom, J. & Segal, E. Contribution of histone sequence preferences to nucleosome organization: proposed definitions and methodology. *Genome Biol.* **11**, 140 (2010).
14. Valouev, A. et al. Determinants of nucleosome organization in primary human cells. *Nature* **474**, 516–520 (2011).
15. Zhang, Z. et al. A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science* **332**, 977–980 (2011).
16. Brogaard, K., Xi, L., Wang, J. P. & Widom, J. A map of nucleosome positions in yeast at base-pair resolution. *Nature* **486**, 496–501 (2012).
17. Valouev, A. et al. A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res.* **18**, 1051–1063 (2008).
18. Hughes, A. L., Jin, Y., Rando, O. J. & Struhl, K. A functional evolutionary approach to identify determinants of nucleosome positioning: a unifying model for establishing the genome-wide pattern. *Mol. Cell* **48**, 5–15 (2012).
19. Teif, V. B. et al. Genome-wide nucleosome positioning during embryonic stem cell development. *Nat. Struct. Mol. Biol.* **19**, 1185–1192 (2012).
20. Szerlong, H. J. & Hansen, J. C. Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure. *Biochem. Cell Biol.* **89**, 24–34 (2011).
21. Ito, T., Bulger, M., Pazin, M. J., Kobayashi, R. & Kadonaga, J. T. ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell* **90**, 145–155 (1997).
22. Fan, Y. et al. Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell* **123**, 1199–1212 (2005).
23. Öberg, C., Izzo, A., Schneider, R., Wrange, Ö. & Belikov, S. Linker histone subtypes differ in their effect on nucleosomal spacing in vivo. *J. Mol. Biol.* **419**, 183–197 (2012).
24. Hartley, P. D. & Madhani, H. D. Mechanisms that specify promoter nucleosome location and identity. *Cell* **137**, 445–458 (2009).
25. Hu, G. et al. Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1. *Genome Res.* **21**, 1650–1658 (2011).
26. Yue, F. et al. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355–364 (2014).
27. Maurano, M. T. et al. Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. *Nat. Genet.* **47**, 1393–1401 (2015).
28. Tolstorukov, M. Y., Volfovsky, N., Stephens, R. M. & Park, P. J. Impact of chromatin structure on sequence variability in the human genome. *Nat. Struct. Mol. Biol.* **18**, 510–515 (2011).
29. Rotem, A. et al. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* **33**, 1165–1172 (2015).

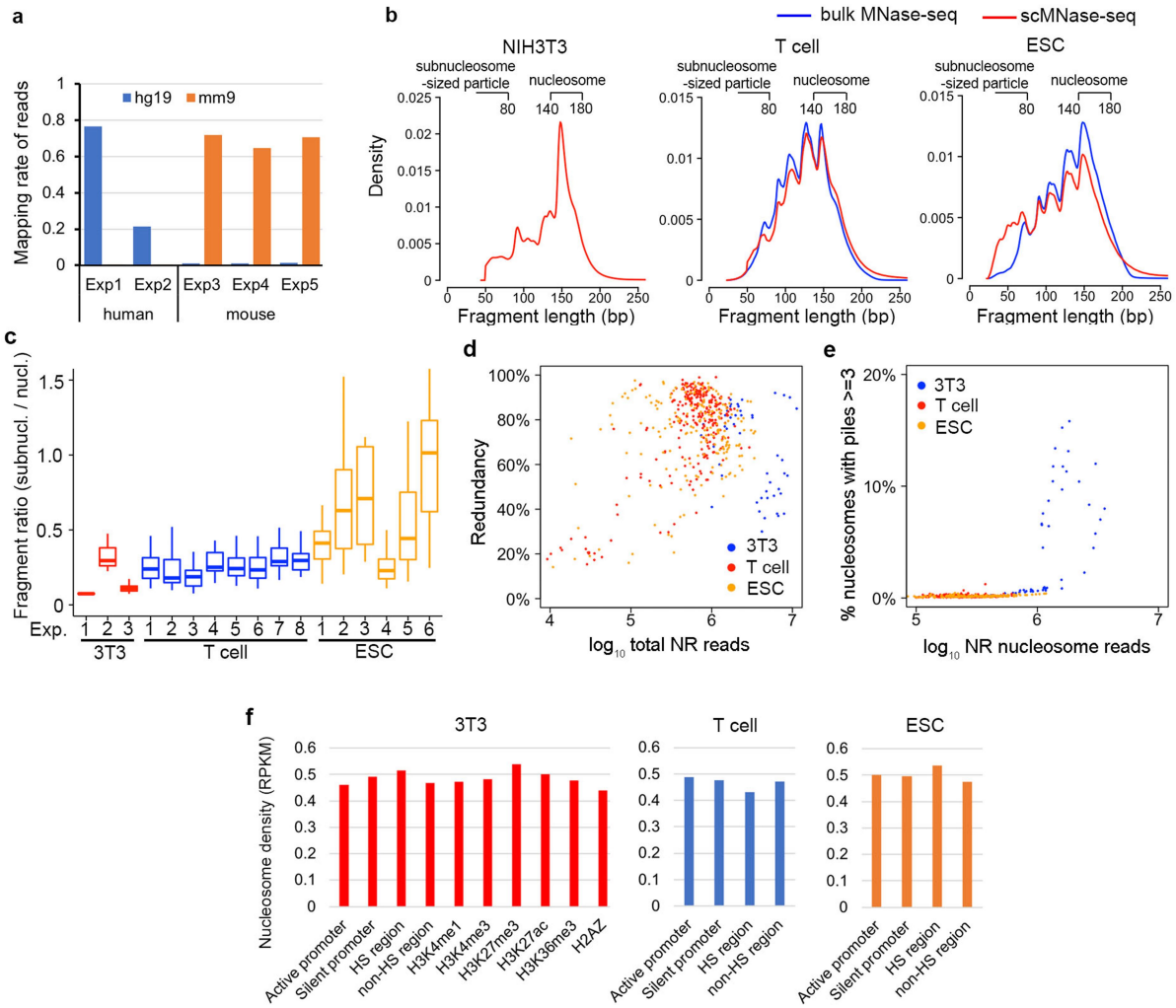
Acknowledgements We thank B. Z. Stanton for critical reading of the manuscript, B. Z. Stanton and J. Cooper for discussions, the National Heart, Lung, and Blood Institute DNA Sequencing Core Facility for sequencing the libraries, the National Heart, Lung, and Blood Institute Flow Cytometry Core facility for sorting the cells, and the National Institutes of Health Biowulf High Performance Computing Systems for computing service. The work was supported by Division of Intramural Research, National Heart, Lung and Blood Institute.

Reviewer information *Nature* thanks M. Buck and the other anonymous reviewer(s) for their contribution to the peer review of this work.

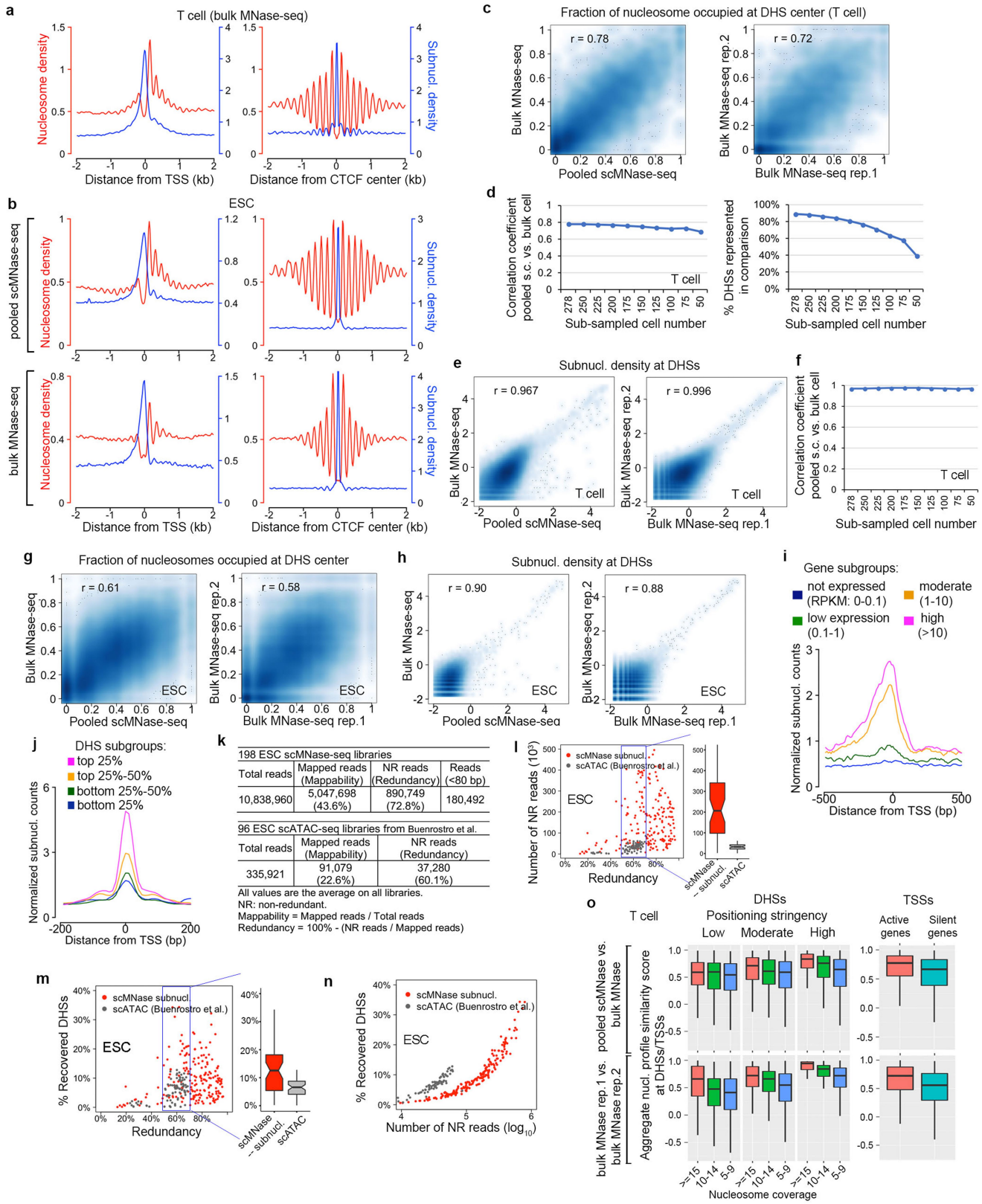
Author contributions K.Z. conceived the project. Q.T. and K.Z. performed the scMNase-seq experiments for NIH3T3 cells. K.C., W.G. and W.X. performed the scMNase-seq experiments for naive T cells and mouse ESCs. B.L. analysed the data. W.J., G.H. and B.N. contributed to the data analysis. B.L. and K.Z. wrote the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0567-3.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0567-3.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to K.Z.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

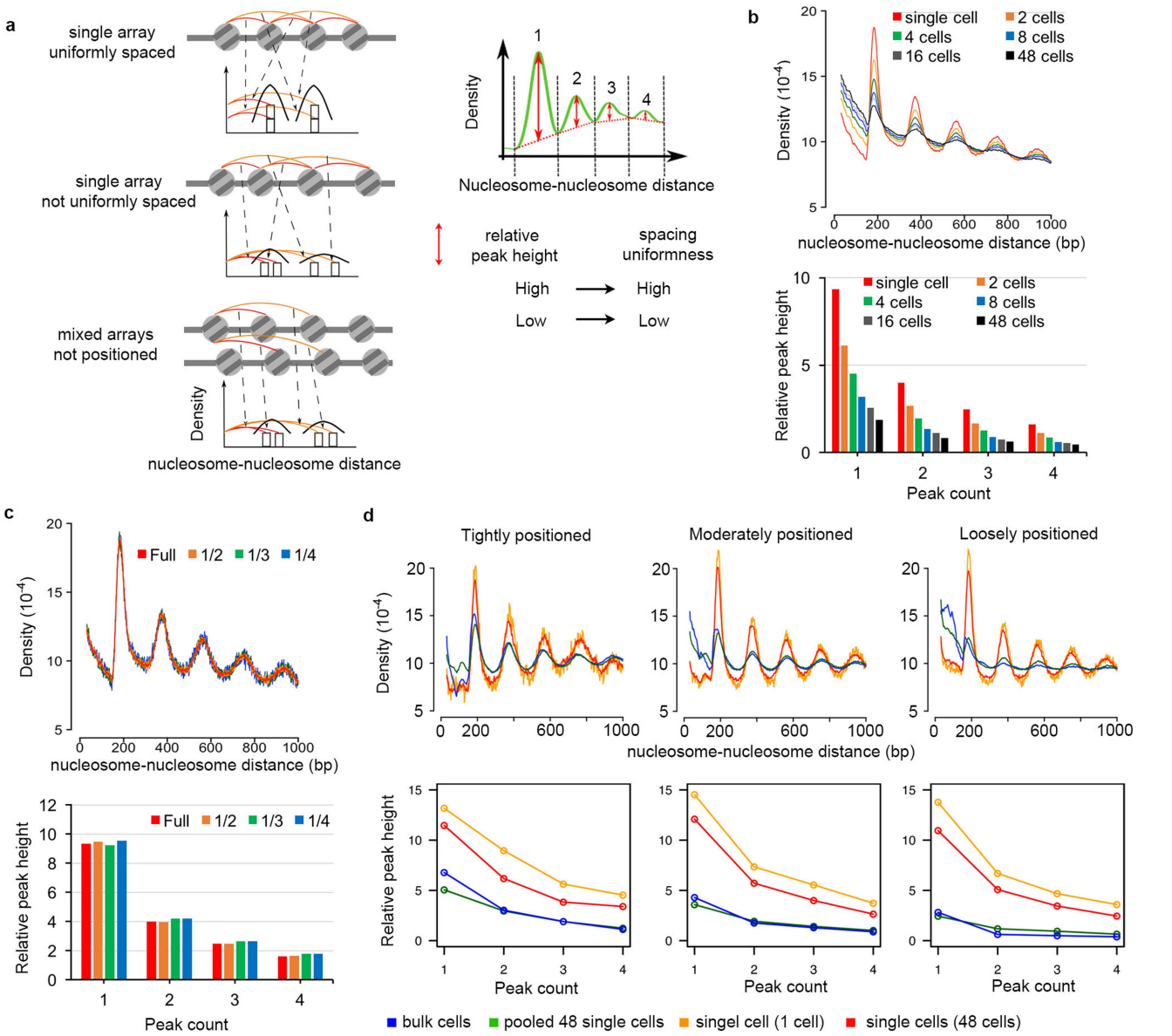


Extended Data Fig. 1 | Characterizing scMNase-seq datasets. **a**, Mapping rates of reads from 100 human cells (two experiments on the left) or 100 mouse cells (three experiments on the right) against human genome (blue) and mouse genome (orange) are shown. The cells were sorted from pre-mixed and MNase-digested human and mouse cells. These data show that there is little contamination of DNA of one cell from another cell. **b**, Fragment-length density of pooled scMNase-seq for NIH3T3 cells, pooled scMNase-seq and bulk-cell MNase-seq for T cells and mouse ESCs. **c**, Box plots of fragment ratio (subnucleosome-sized particle-to-nucleosome) for NIH3T3 cell, naive CD4 T cell and mouse ESC scMNase-seq libraries. Single-cell libraries were grouped by biologically independent experiments. Supplementary Table 1 gives the library number for each group. Centre line, median; boxes, first and third quartiles; whiskers, 1.5× interquartile range. **d**, Plot of non-redundant (NR) read number (*x* axis) and sequencing redundancy (*y* axis) for single NIH3T3 cells, CD4 T cells and mouse ESCs. **e**, Plot of non-redundant nucleosome reads (*x* axis) and percentage of nucleosomes with overlapping piles ≥ 3 (*y* axis). The plot suggests the polyploidy of NIH3T3 cells. **f**, Nucleosome density at different genomic regions for NIH3T3 cell, CD4 T cell and mouse ESC scMNase-seq libraries reveals that the nucleosomes in different genomic regions were similarly detected and represented by scMNase-seq.

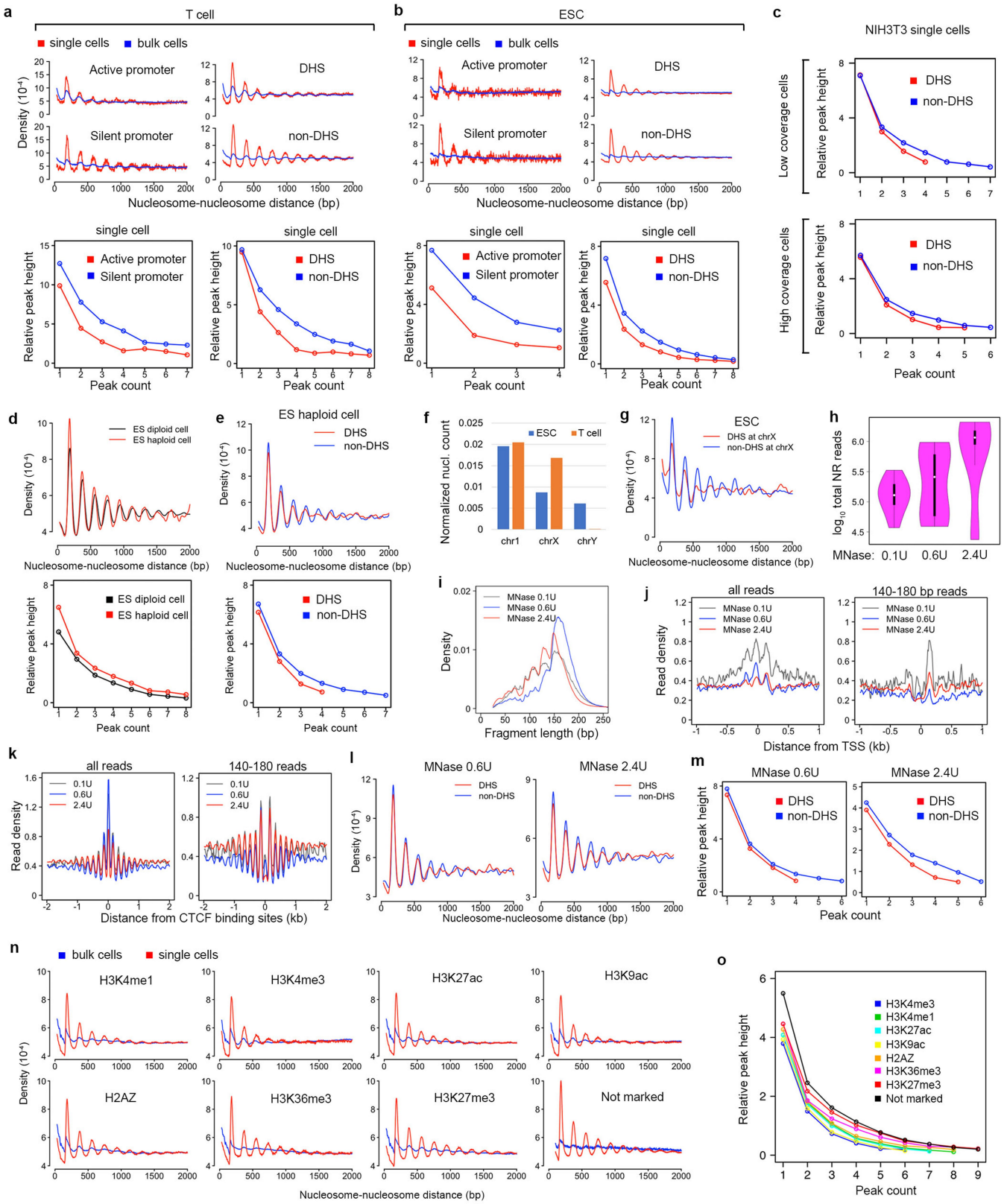


Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Characterizing pooled scMNase-seq data and subnucleosome-sized particles. a, Average density profiles of nucleosomes (red) and subnucleosome-sized particles (blue) relative to the TSSs of active genes (left) and CTCF-binding sites (right) for bulk-cell naive CD4 T cell MNase-seq data. **b**, Average density profiles of nucleosomes (red) and subnucleosome-sized particles (blue) relative to the TSSs of active genes (left) and CTCF-binding sites (right) for pooled mouse ESC scMNase-seq data (top) and bulk-cell mouse ESC MNase-seq data (bottom). **c**, Smoothed scatter plot for the fraction of nucleosome occupied at 8,929 DHS centres (selected from the top 10,000 DHSs (see Supplementary Methods for criteria)) for pooled scMNase-seq (*x* axis) versus bulk-cell MNase-seq (*y* axis) for T cells (left). Pearson correlation coefficient is indicated. As a positive control, the scatter plots for two bulk-cell MNase-seq replicates are also shown (right). **d**, Pearson correlation coefficient for the fraction of nucleosomes occupied at the DHS centre between pooled sub-sampled CD4 T cell scMNase-seq libraries and bulk-cell MNase-seq, as a function of sub-sampled cell number (left). Percentages of top 10,000 DHSs represented in the comparison—that is, the sample size in the top panel—as a function of sub-sampled cell number are also shown (right). **e**, Smoothed scatter plot for subnucleosome-sized particle density at 83,229 DHSs for pooled scMNase-seq (*x* axis) versus bulk-cell MNase-seq (*y* axis) for T cells (left). Pearson correlation coefficient is indicated. As a positive control, the scatter plots for two bulk-cell MNase-seq replicates are also shown (right). **f**, Pearson correlation coefficient for subnucleosome-sized particle density between pooled sub-sampled T cell scMNase-seq libraries and bulk-cell MNase-seq at 83,229 DHSs, as a function of sub-sampled cell number. **g**, **h**, Smoothed scatter plot for the fraction of nucleosomes occupied at 8,449 DHS centres (selected from top-10,000 DHSs (see Supplementary Methods for criteria)) (**g**) and subnucleosome-sized particle density at 94,250 DHSs (**h**) for pooled scMNase-seq (*x* axis) versus bulk-cell MNase-seq (*y* axis) for mouse ESCs. Pearson correlation coefficient is indicated. As a positive control, the scatter plots for two bulk-cell MNase-seq replicates are also shown. **i**, **j**, Average density profiles of subnucleosome-sized particles around TSSs for gene subgroups with different expression levels (**i**) and around DHSs for DHS subgroups with different DNase I tag densities (**j**). **k**, Table showing the mapping statistics for 198 mouse ESC scMNase-seq libraries and 96 previously published[10] mouse ESC scATAC-seq libraries. **l**, **m**, Scatter plots of the number of non-redundant reads (**l**, *y* axis) and percentage of recovered DHSs (**m**, *y* axis) versus sequencing redundancy (*x* axis) for scMNase-seq subnucleosome-sized particles (red, $n = 198$ single-cell libraries) and scATAC-seq reads (grey, $n = 96$ single-cell libraries). Box plots (right) show the values from scatter plots (left) for cells with redundancy that ranges from 50% to 70% (blue rectangle in the left panel; red, $n = 49$; grey, $n = 58$) for the two methods. **n**, Scatter plot showing the percentage of recovered DHSs (*y* axis) versus number of non-redundant reads for scMNas-seq subnucleosome-sized particles (red, $n = 198$ single-cell libraries) and scATAC-seq reads (grey, $n = 87$ single-cell libraries). **o**, Aggregated nucleosome profile similarity score at DHSs for different groups of DHSs (left) and two promoter groups (right), for comparison between pooled scMNase-seq and bulk-cell MNase-seq (top) and between two bulk-cell MNase-seq replicates (bottom). The DHS groups are classified by three positioning-stringency levels (low, positioning score < 0.45; moderate, $0.45 <$ positioning score < 0.65, high, positioning score > 0.65) and three nucleosome coverage levels (high, ≥ 15; moderate, 10–15; low, 5–9). The DHS numbers for each group are: low positioning score and high coverage, $n = 803$; low positioning score and moderate coverage, $n = 531$; low positioning score and low coverage, $n = 450$; moderate positioning score and high coverage, $n = 701$; moderate positioning score and moderate coverage, $n = 592$; moderate positioning score and low coverage, $n = 588$; high positioning score and high coverage, $n = 162$; high positioning score and moderate coverage, $n = 230$; high positioning score and low coverage, $n = 395$. The number of promoters for each group: active, $n = 6{,}777$; silent, $n = 418$. In box plots in **i**, **m**, **o**, centre line, median; boxes, first and third quartiles; whiskers, 1.5× interquartile range; notch, 95% confidence interval of the median.

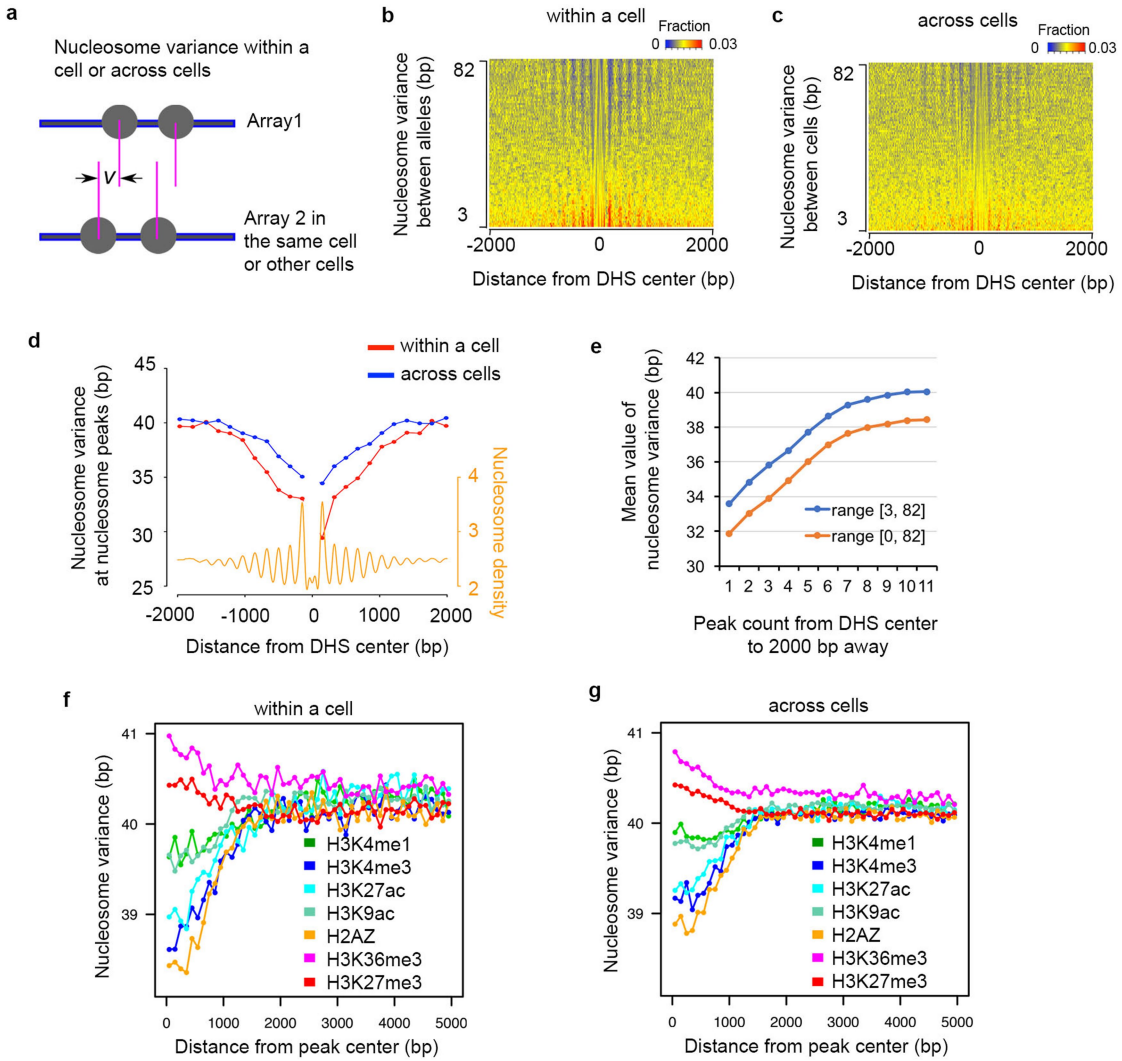


Extended Data Fig. 3 | Measuring uniformity in nucleosome spacing in single cells. a, Cartoon illustrates that uniformity in nucleosome spacing can be measured by nucleosome-to-nucleosome distance density: uniformly spaced nucleosomes in a single array result in sharp and high peaks, whereas non-uniformly spaced nucleosomes result in flat peaks or no peaks. Nucleosomes from mixed arrays also result in flat peaks, even if they are uniformly spaced. **b**, The nucleosome space phasing and relative peak height gradually decreased as the number of mixed cells increases, which indicates cellular heterogeneity of nucleosome organization across different cells. **c**, Nucleosome space phasing and relative peak height do not change when reducing the library size (number of sequence reads) to 1/2, 1/3 and 1/4. **d**, Density plots of nucleosome-to-nucleosome distance (top) and relative peak height in density plot (bottom) for nucleosomes with different positioning stringency for bulk-cell MNase-seq, pooled 48 single cells, one representative single cell and 48 single-cell scMNase-seq datasets.

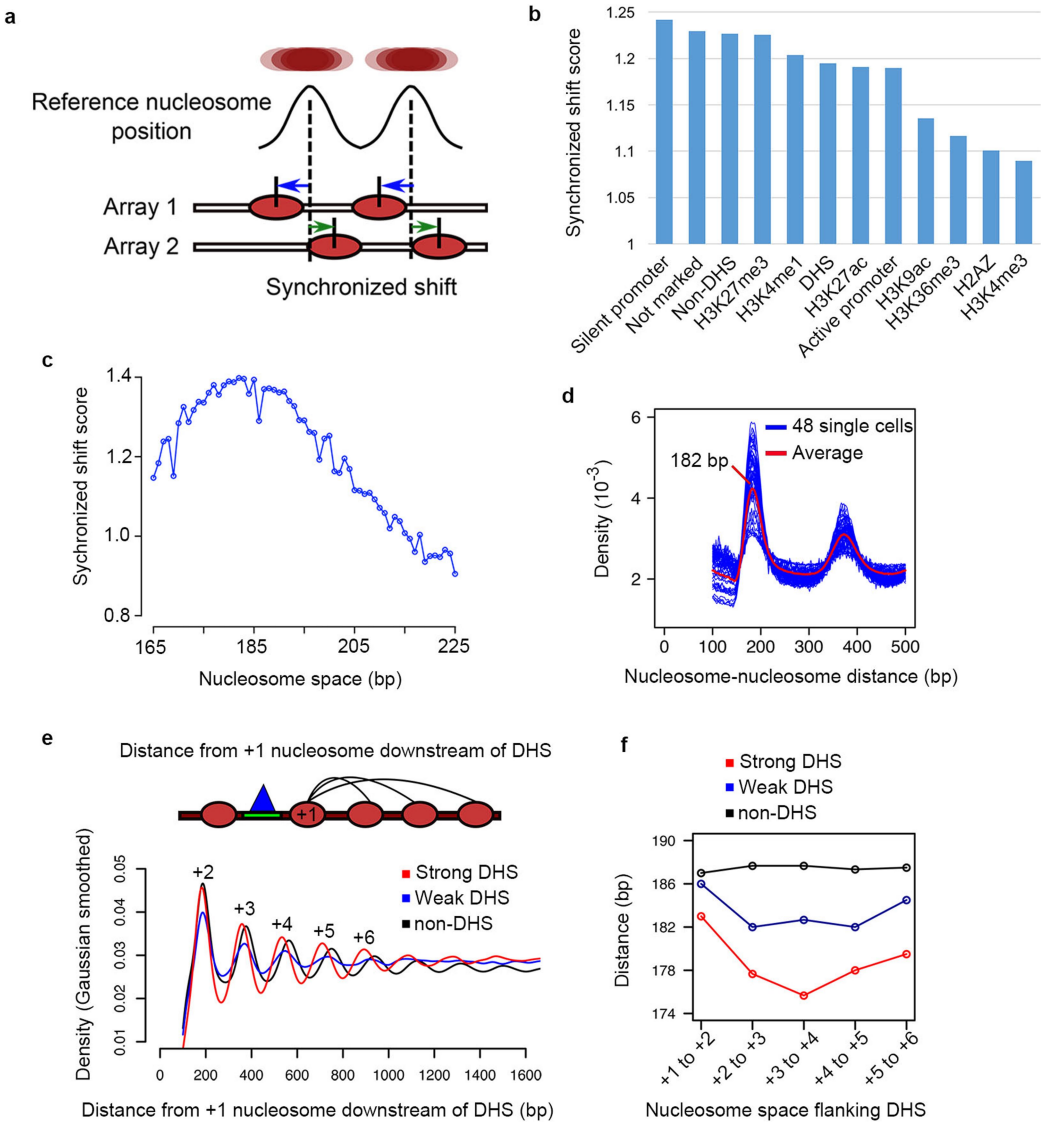


Extended Data Fig. 4 | See next page for caption.

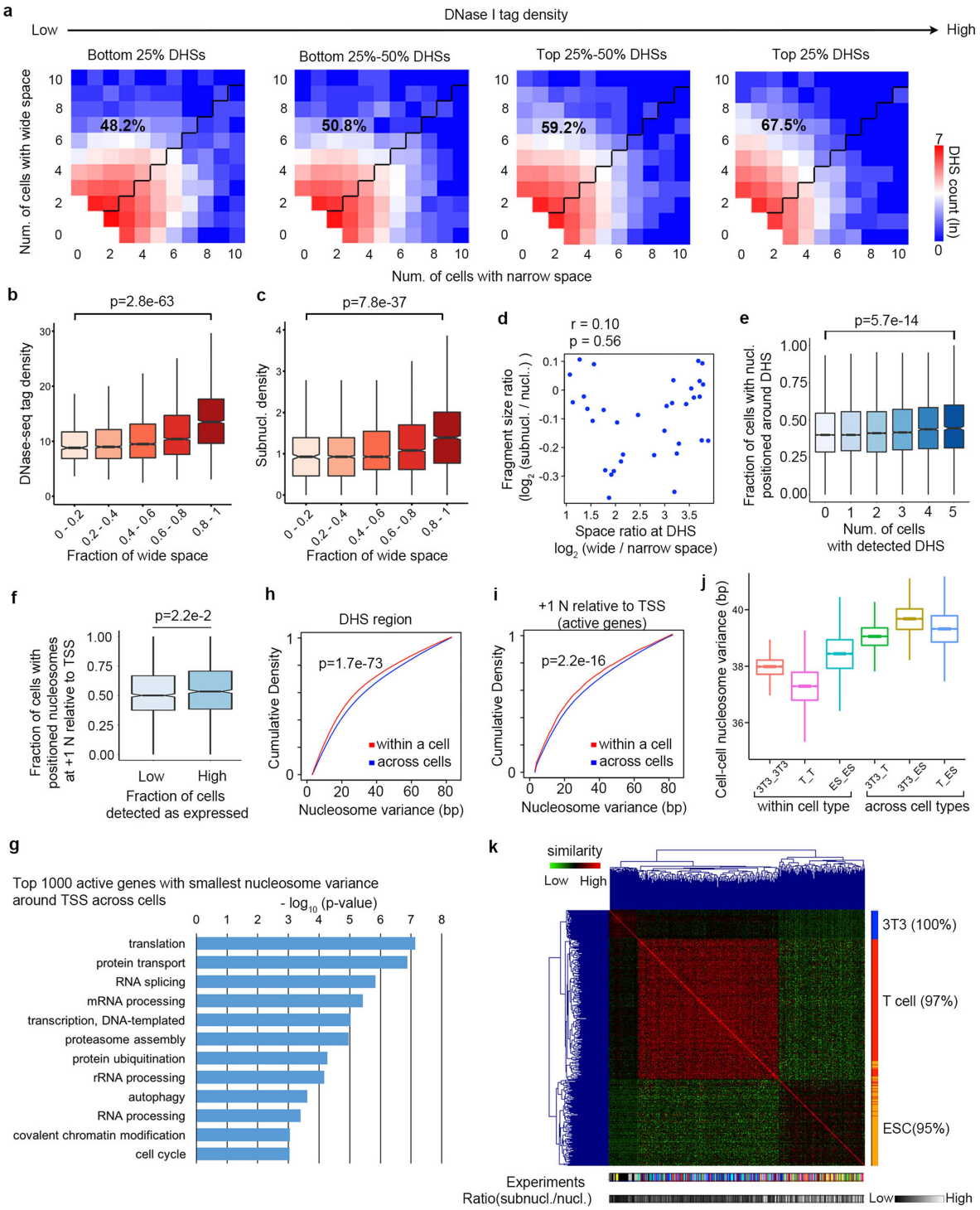
Extended Data Fig. 4 | Uniformity in nucleosome spacing is higher in silent heterochromatin regions than in active chromatin regions. **a**, **b**, Density plots of nucleosome-to-nucleosome distance (top) and relative peak height (bottom) for nucleosomes at active or silent promoters and DHS or non-DHS regions for T cells (**a**) and mouse ESCs (**b**). **c**, Relative peak height of density plots for nucleosome-to-nucleosome distance for nucleosomes in DHS (red) and non-DHS regions (blue) for low-coverage cells (top) and high-coverage cells (bottom). **d**, Density plots of nucleosome-to-nucleosome distance (top) and relative peak height (bottom) for diploid (black) and haploid (red) mouse ESCs. **e**, Density plots of nucleosome-to-nucleosome distance (top) and relative peak height (bottom) at DHS and non-DHS regions for haploid mouse ESCs. **f**, Mapped nucleosome count normalized by chromosome length at chromosome 1, X and Y for mouse ESCs and CD4 T cells suggests that mouse ESCs are derived from male mouse. **g**, Density plots of nucleosome-to-nucleosome distance at DHS and non-DHS regions at chromosome X for mouse ESCs. **h**, Violin plots of library size (total non-redundant reads) for NIH3T3 scMNase-seq libraries treated with three MNase concentrations (0.1 unit (0.1 U), 0.6 unit (0.6 U) and 2.4 unit (2.4 U) MNase per million cells). Each condition has 10 single-cell libraries. In the violin plots, centre dot, mean; inner layer, the interquartile range. **i**, Fragment-length density of pooled scMNase-seq data with three MNase concentrations. **j**, **k**, Average density profiles of all reads (**j**, **k**, left) and nucleosome reads with length between 140 and 180 bp (**j**, **k**, right) around the TSSs of active genes (**j**) and CTCF-binding sites (**k**) for pooled scMNase-seq with three MNase concentrations. **l**, **m**, Density plots of nucleosome-to-nucleosome distance (**l**) and relative peak height (**m**) at DHS (red) and non-DHS (blue) regions for scMNase-seq treated by 0.6 U (left) and 2.4 U (right) MNase concentrations. **n**, **o**, Density plots of nucleosome-to-nucleosome distance (**n**) and relative peak height (**o**) for nucleosomes within genomic regions marked by different histone modifications.



Extended Data Fig. 5 | Variation in nucleosome positioning within a cell or across different single cells around the centre of DHS or histone modification peaks. a, Cartoon illustrating the definition of nucleosome variance within a cell or across different single cells. **b**, **c**, Heat maps showing the distribution of nucleosome variance at the position relative to DHS centre within a cell (**b**) or across different single cells (**c**). **d**, Nucleosome variance within a cell (red) and across single cells (blue) becomes smaller when getting closer to the DHS centre. **e**, Calculations of mean value of nucleosome variance from two ranges (3–82 bp and 0–82 bp) reveal the same trend of increase when nucleosomes become farther away from DHS centre. **f**, **g**, Average profiles of nucleosome variance at the position relative to the centre of histone modification peaks within a cell (**f**) or across different cells (**g**).

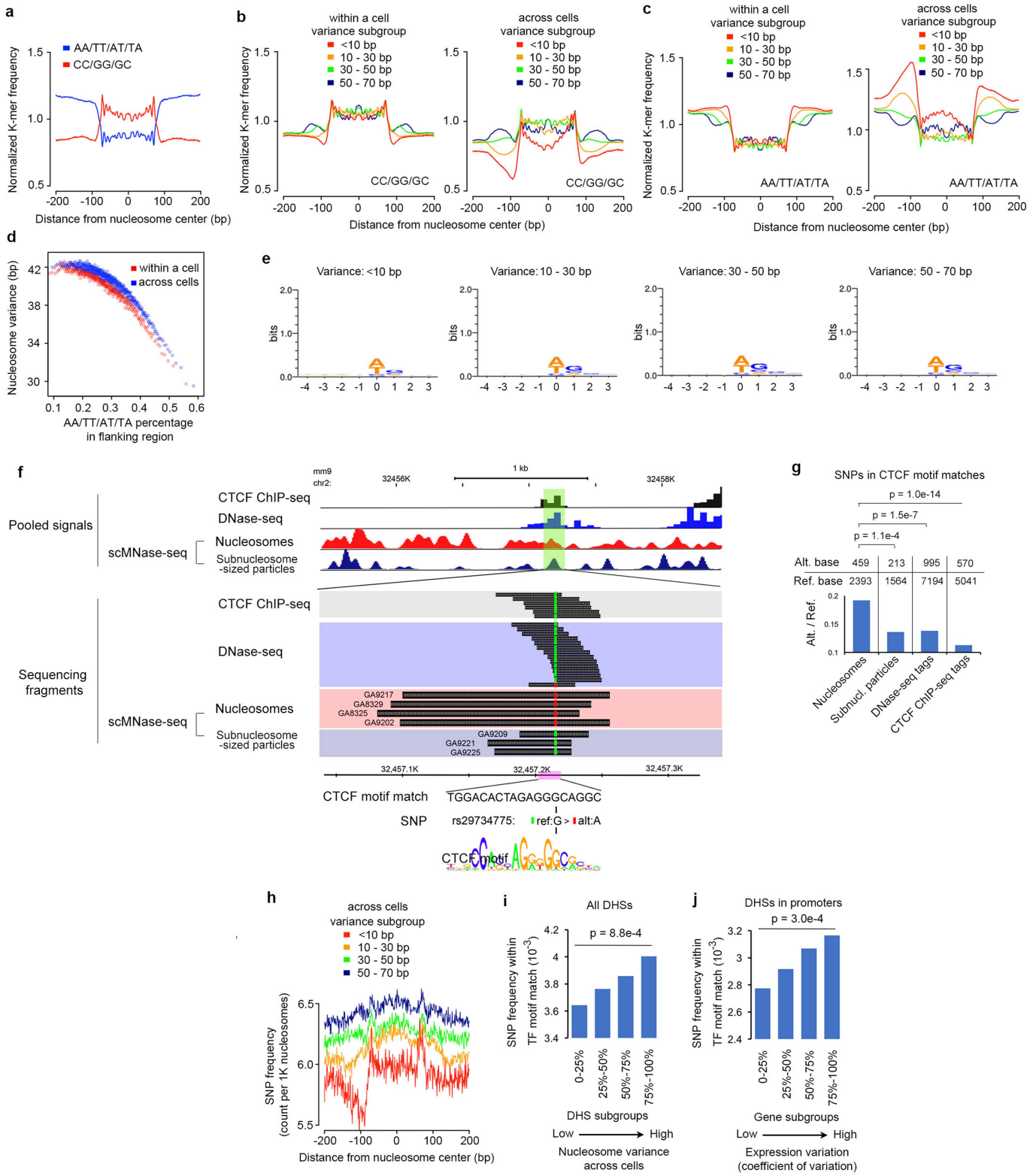


Extended Data Fig. 6 | Nucleosomes show a synchronized shift in silent-gene promoters and heterochromatin regions, and show compressed spacing where they flank DHS centres. a, Cartoon illustrates synchronized shift of adjacent nucleosomes within single nucleosome arrays. **b**, Bar plot showing synchronized shift score for different genomic regions. Silent promoter, silent-gene promoter; active promoter, active-gene promoter; not marked, regions not marked by any histone modifications as shown; DHS: ±2,000-bp region surrounding DHS centre; non-DHS, intervals of DHS regions. **c**, Synchronized shift score for nucleosome pairs with different distances of nucleosome space. **d**, Density plot of nucleosome-to-nucleosome distance in single cells reveals dominant nucleosome space at about 182 bp. **e**, Density plot of nucleosome spacing in the regions flanking strong and weak DHSs as well as non-DHSs. **f**, Distances between each pair of nucleosomes in the chromatin regions flanking strong DHS, weak DHS or non-DHS, described in **e**.



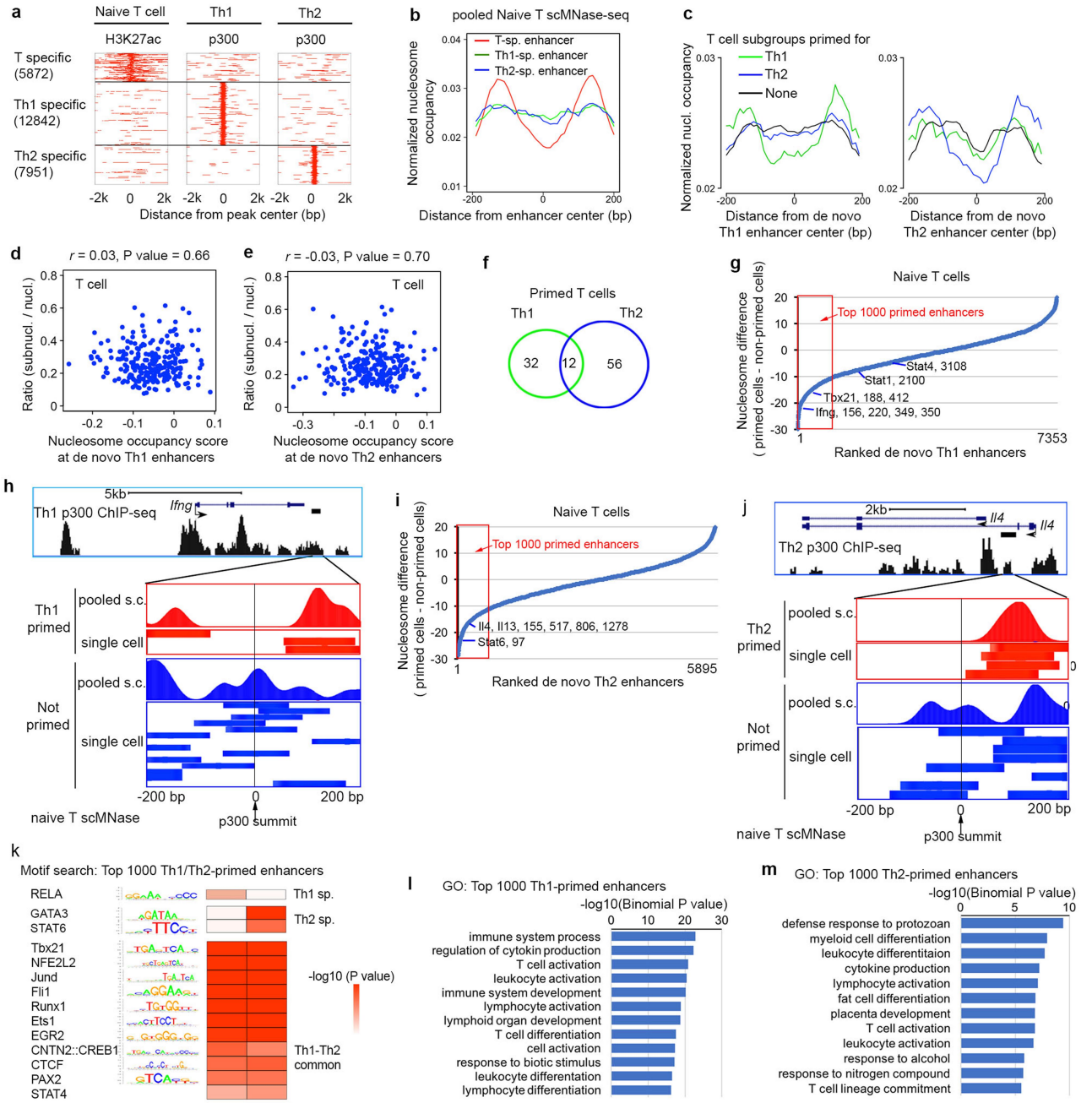
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Heterogeneity of nucleosome spacing and positioning around DHS across different single cells. a, Heat maps showing DHS frequency as a function of number of cells with the narrow spacing (*x* axis) and number of cells with the wide spacing (*y* axis) for four DHS subgroups with different DNase I tag densities. Numbers indicate the percentages of DHSs that have more wide space than narrow space. **b**, **c**, Box plots showing the accessibility level from cell population, measured by DNase-seq tag density (**b**) and pooled scMNase-seq subnucleosome-sized particle density (**c**), for five groups of DHSs defined by fraction of wide space. Data represent values on 612, 2,088, 3,858, 2,500 and 1,586 DHSs (from left to right). **d**, Scatter plot of the ratio of wide-to-narrow space at the DHS in a single cell (*x* axis) and fragment-size ratio of subnucleosome-sized particles to nucleosomes (*y* axis) on 48 NIH3T3 scMNase-seq libraries. Pearson correlation coefficient and *P* value are indicated. *P* value is the probability that one would have found the current result if the correlation coefficient were zero (null hypothesis), and was calculated using R. **e**, Box plot showing fraction of cells with positioned nucleosomes around a DHS for different groups of DHSs. DHSs were grouped on the basis of the number of cells detected as DHS in a previously published scDNase-seq experiment[12]. Number of DHSs for each group was 44,040, 15,622, 11,056, 8,009, 4,063 and 1,180 (from left to right). **f**, Box plot showing fraction of cells with a positioned +1 nucleosome for two groups of genes sorted by expression variation (low, $n = 1{,}171$; high, $n = 1{,}174$). **g**, Gene Ontology analysis of top-1,000 active genes with the smallest nucleosome variance across cells. Significant Gene Ontology terms with *P* value are reported by David Bioinformatics Resources (v.6.7). **h**, Density plot showing nucleosome variance around DHSs within a cell ($n = 73{,}274$ nucleosome pairs) and across different cells ($n = 752{,}398$ nucleosome pairs). **i**, Cumulative density plot for nucleosome variation at +1 nucleosome relative to the TSSs of active genes within a cell (red, $n = 11{,}388$ nucleosome pairs) and across cells (blue, $n = 237{,}006$). **j**, Box plot showing nucleosome variance around DHSs across cells for within a cell type (NIH3T3–NIH3T3 cells, $n = 1{,}128$ nucleosome pairs; T cells–T cells, $n = 23{,}936$; ESCs–ESCs, $n = 5{,}775$) and across different cell types (NIH3T3–T cells, $n = 11{,}856$; NIH3T3–ESCs, $n = 6{,}962$; T cells–ESCs, $n = 20{,}442$). **k**, Heat map reveals clustering of NIH3T3 cells, T cells and mouse ESCs based on cell-to-cell nucleosome dissimilarity score around DHSs. Colour bar on the right indicates cell types and colour bars on the bottom indicate experiment time and fragment-size ratio. *P* values in panels **b**, **c**, **e**, **f**, **h** and **i** were calculated using one-sided Mann–Whitney *U*-test. In box plots, centre line, median; boxes, first and third quartiles; whiskers, 1.5× interquartile range; notch, 95% confidence interval of the median.

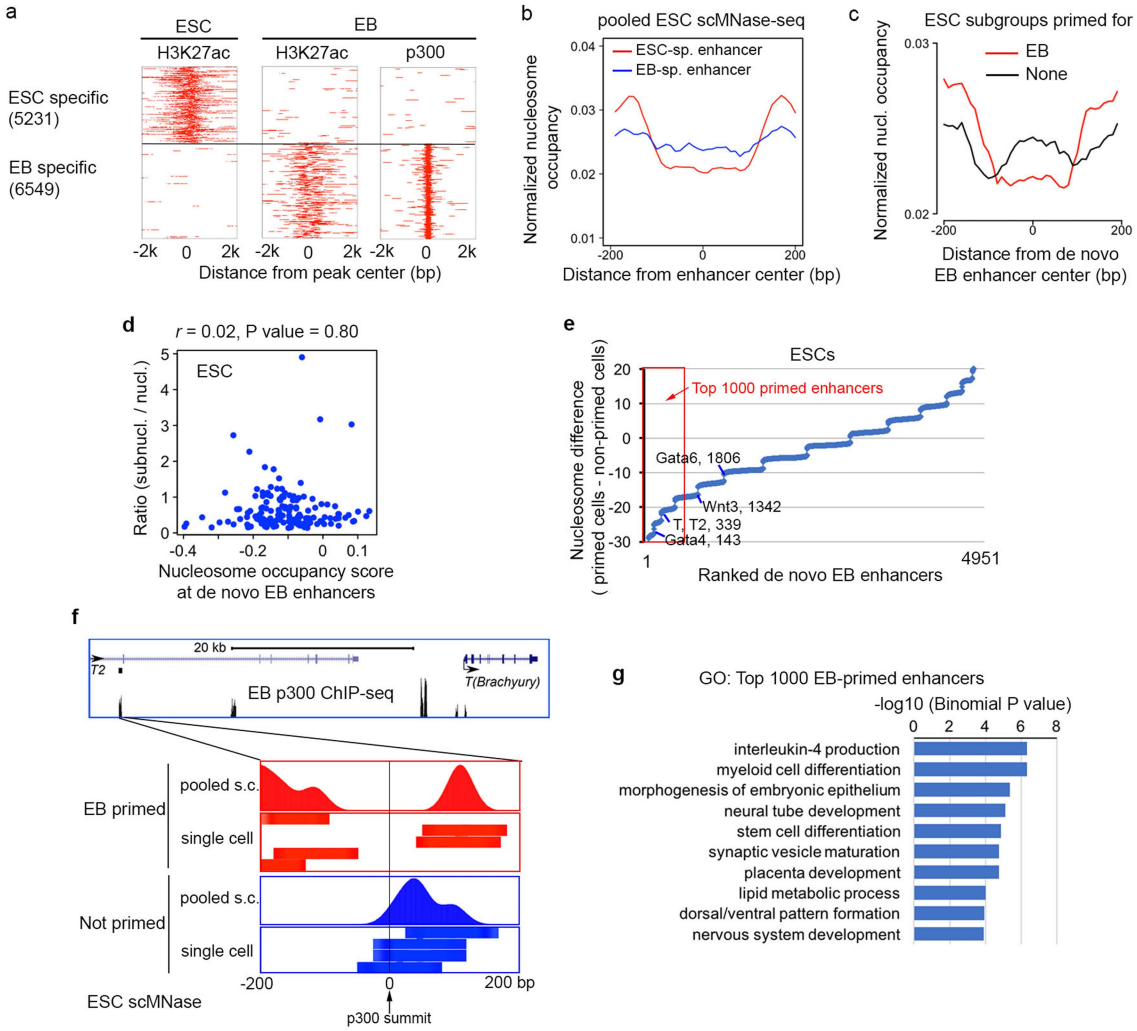


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Cell-to-cell single-base variation is associated with variation in nucleosome positioning and variation in gene expression across different single cells. a, CC, GG and GC frequency is higher in the nucleosome-occupied region than in the flanking region, whereas AA, TT, AT and TA frequency shows the opposite pattern. **b**, CC, GG and GC frequency in flanking regions increases as nucleosome variance within a cell (left) or across different single cells (right) increases. **c**, AA, TT, AT and TA frequency in flanking regions decreases as nucleosome variance within a cell (left) or across different single cells (right) increases. **d**, Nucleosome variances within a cell and across different single cells are reversely correlated with the percentage of AA, TT, AT and TA in flanking regions. **e**, Weblogos sequences logos for sequence preferences across MNase cleavage sites are shown for subgroups of nucleosomes with different positioning variance across cells. **f**, An example showing a CTCF motif with the reference base (green) in some cells and alternative base (red) in other cells. scMNase-seq data show that the reference base is associated with subnucleosome-sized particles, whereas the alternative base is associated with the nucleosome structure. Fragments from DNase-seq and CTCF ChIP-seq datasets within the window are also shown with the bases at single-nucleotide polymorphism location highlighted. Tracks for tag densities of CTCF ChIP-seq, DNase-seq, and nucleosomes and subnucleosome-sized particles from pooled single cells are shown in a zoomed-out window. **g**, The number of CTCF-motif matches containing alternative or reference bases at the single-nucleotide polymorphism locus occupied by nucleosomes, subnucleosome-sized particles, sequence reads obtained by DNase-seq and by CTCF ChIP-seq. *P* value was calculated using one-sided Fisher's exact test. The ratio between alternative and reference bases is also shown (bottom). **h**, Single-nucleotide polymorphism frequency is correlated with nucleosome variation across different single cells. Variant frequencies at each position relative to nucleosome midpoint for four nucleosome subgroups with different levels of nucleosome variance across cells are shown. **i**, Single-nucleotide polymorphism frequency within transcription-factor motifs at DHSs for four DHS subgroups, sorted by nucleosome variance around DHS across different single cells (each subgroup has 22,139 DHSs that contains at least one transcription-factor motif match). **j**, Single-nucleotide polymorphism frequency within transcription-factor motifs at DHSs in promoters for gene subgroups, sorted by expression variation across different single cells (each subgroup has 2,136 genes). *P* value in **i**, **j** is defined as the probability of observing a larger difference than current result between two groups by random. *P* value calculation is described in Supplementary Methods. SNP, single-nucleotide polymorphism.



Extended Data Fig. 9 | Characterization of primed enhancers in undifferentiated naive CD4 T cells. a, Heat maps show H3K27ac in naive T cells and p300 in T_H1 and T_H2 cells around naive T cell-specific, T_H1-specific and T_H2-specific enhancers. **b**, Profile of nucleosome occupancy from pooled naive T cell scMNase-seq around T cell-specific, T_H1-specific and T_H2-specific enhancers. **c**, Normalized nucleosome occupancy within ± 200 bp of the centre of de novo T_H1 enhancers (left) or de novo T_H2 enhancers (right) for subgroups of T cells primed for T_H1 cells (green), T_H2 cells (blue) or none (black). **d**, **e**, Plots of fragment-size ratio of subnucleosome-sized particles-to-nucleosomes versus nucleosome occupancy score at de novo T_H1 (**d**) and T_H2 (**e**) enhancers for 237 naive CD4 T cells reveal that nucleosome occupancy score is not correlated with fragment-size ratio. Pearson correlation coefficient and *P* value are indicated. *P* value is the probability that one would have found the current result if the correlation coefficient were zero (null hypothesis), and was calculated using R. **f**, Subgroups of naive CD4 T cells primed for T_H1 and T_H2 do not have much overlap. **g**, Plots of de novo T_H1 enhancers ranked on the basis of differences in nucleosome occupancy between pooled primed cells and the non-primed cells (*y* axis, see Supplementary Methods). Enhancers associated with key genes for T_H1 were labelled by genes along with ranks. **h**, Nucleosome positions in pooled or single primed (red) and non-primed (blue) cells at de novo T_H1-specific enhancers for *Ifng* gene. **i**, Plots of de novo T_H2 enhancers ranked on the basis of differences in nucleosome occupancy between pooled primed cells and the non-primed cells (*y* axis, see Supplementary Methods). Enhancers associated with key genes for T_H2 were labelled by genes along with ranks. **j**, Nucleosome positions in pooled or single primed (red) and non-primed (blue) cells at de novo T_H2-specific enhancers for *Il4* gene. **k**, Motifs enriched in the top 1,000 T_H1/T_H2-primed enhancers are shown. **l**, **m**, Gene Ontology analysis for top 1000 T_H1-primed (**l**) and T_H2-primed (**m**) enhancers. Significant Gene Ontology terms with P values are reported using GREAT v.3.0.0.



Extended Data Fig. 10 | Characterization of primed enhancers in undifferentiated mouse ESCs. a, Heat maps show H3K27ac in mouse ESCs and EB cells and p300 in EB cells around ESC-specific and EB-specific enhancers. **b**, Profile of nucleosome occupancy from pooled mouse ESC scMNase-seq around mouse ESC-specific and EB-specific enhancers. **c**, Normalized nucleosome occupancy within ± 200 bp of the centre of de novo EB-specific enhancers for subgroups of mouse ESCs that are primed for EB (red) or not primed for EB (black). **d**, Plots of fragment-size ratio of subnucleosome-sized particles-to-nucleosomes versus nucleosome occupancy score at de novo EB-specific enhancers for 144 mouse ESCs. Pearson correlation coefficient and *P* value are indicated. *P* value is the probability that one would have found the current result if the correlation coefficient were zero (null hypothesis), and was calculated using R. **e**, Plots of de novo EB-specific enhancers ranked on the basis of difference in nucleosome occupancy between pooled primed cells and the non-primed cells. Enhancers associated with key genes for EB cells were labelled by genes along with ranks. **f**, Nucleosome positions in pooled or single primed (red) and non-primed (blue) cells at de novo EB-specific enhancers for *Brachyury* gene. **g**, Gene Ontology analysis for top-1,000 EB-primed enhancers. Significant Gene Ontology (GO) terms with *P* values are reported using GREAT v.3.0.0.

nature research

Corresponding author(s): Keji Zhao

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☒	☐	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	Bowtie 2 (v2.2.5) was used for sequencing reads mapping. MACS2 (v2.1.0) was used for ChIP-seq peaks calling. Cistrome SeqPos was used to do motif analysis. PIQ (v1.3) was used to predict TF binding sites in DHSs. Other data analysis, e.g. nucleosome spacing uniformness quantification and nucleosome occupancy score, was performed using custom codes and fully described in Methods. Custom codes for nucleosome spacing uniformness quantification and nucleosome occupancy score are available at https://github.com/binbinlai2012/scMNase.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The scMNase-seq data sets have been deposited in the Gene Expression Omnibus database with accession number GSE96688.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We generated scMNase-seq datasets for 48 NIH3T3 single cells, 198 mouse embryonic stem cells (mESCs), and 278 mouse naïve CD4 T cells. In addition, we generated three groups of NIH3T3 scMNase-seq libraries (10 single cells for each group) with different MNase concentrations for titration purpose, and 12 haploid mESC scMNase-seq libraries. Statistical tests were not used to pre-determine sample size. ~500 single cell libraries covering 3 cell types generates total 500 million high quality non-redundant sequencing reads used for determining nucleosome position, subnucleosome-sized particle position and nucleosome-nucleosome distance from the same cells or across cells, within or across cell types. This data set fully supports the findings described in the manuscript.
Data exclusions	All data were included for assessing or characterizing the scMNase-seq libraries. Single cells with low library size were excluded for the analysis of "clustering single cells based on nucleosome similarity around DHSs" and the analysis of "identifying subgroup of naive CD4 T cells or mESCs with depleted nucleosomes at de novo enhancers for differentiation". Details were described in the methods section.
Replication	We have three cell types and in total ~500 single cells to support our findings related to the single cell nucleosome profiling. In all three cell types, we observed that nucleosome spacing uniformness is smaller in silent chromatin region than active chromatin region. In both mESCs and naive CD4 T cells, which are undifferentiated cells, we found a subgroup of primed cells show nucleosome depletion at downstream lineage specific de novo enhancers. For the analysis of bimodal nucleosome spacing distribution at DHSs and its link to heterogeneity in chromatin accessibility and gene expression, we only used NIH3T3 cells because (1) only NIH3T3 cell libraries have sufficient nucleosome coverage to provide sufficient nucleosome pairs around DHSs, and (2) only NIH3T3 cells have published scDNase-seq and single cell Drop-seq data sets available for this integrative analysis.
Randomization	For each cell type, single cells were FACS sorted from similar cell samples. There was no randomization in this in vitro experiment.
Blinding	This study is in vitro experiment and no blinding is required. The investigators were not blinded during data collection and analysis.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	NIH3T3 cells and mESCs were obtained from ATCC (ATCC® CRL-1658™ and ATCC® SCRC-1036™).
Authentication	In addition to the authentic commercial source, our own data (RNA-seq and MNase-seq) from these cell lines validate the identity of these cells.
Mycoplasma contamination	The cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.