RESOURCE

nature structural & molecular biology

Genome-wide nucleosome positioning during embryonic stem cell development

Vladimir B Teif[1,2], Yevhen Vainshtein[2,3], Maïwen Caudron-Herger[1,2], Jan-Philipp Mallm[1,2], Caroline Marth[1,2], Thomas Höfer[2,3] & Karsten Rippe[1,2]

We determined genome-wide nucleosome occupancies in mouse embryonic stem cells and their neural progenitor and embryonic fibroblast counterparts to assess features associated with nucleosome positioning during lineage commitment. Cell-type- and protein-specific binding preferences of transcription factors to sites with either low (Myc, Klf4 and Zfx) or high (Nanog, Oct4 and Sox2) nucleosome occupancy as well as complex patterns for CTCF were identified. Nucleosome-depleted regions around transcription start and transcription termination sites were broad and more pronounced for active genes, with distinct patterns for promoters classified according to CpG content or histone methylation marks. Throughout the genome, nucleosome occupancy was correlated with certain histone methylation or acetylation modifications. In addition, the average nucleosome repeat length increased during differentiation by 5–7 base pairs, with local variations for specific regions. Our results reveal regulatory mechanisms of cell differentiation that involve nucleosome repositioning.

Embryonic stem cells and differentiated cells derived from them share the same DNA sequence but have distinct cellular functions. Many of the underlying cell-fate decisions occur through changes to chromatin features that affect gene expression. The specific location of nucleosomes on the DNA is important for controlling access to the DNA[1,2]. Binding of protein factors to the 145–147 base pairs of DNA wrapped around the histone octamer core is frequently impeded, whereas the linker DNA between nucleosomes is more easily accessible. Recent advancements in high-throughput sequencing methods allowed for the genome-wide mapping of individual nucleosomes at single–base pair resolution[3,4], with yeast serving as a model system for the initial pioneering studies[5–7]. More recently, tissue- and disease-specific features of nucleosome positions in higher organisms were reported[8–14]. These include studies of human cell lines[8–10] and mouse hepatocyte cells[15]. We set out to identify features of nucleosome positioning at functional genomic elements during lineage commitment in mouse embryonic stem cells (ESCs) and neural progenitor cells (NPCs) derived from these ESCs, as well as mouse embryonic fibroblasts (MEFs) from the corresponding mouse strain. By comparing these three cell types, we identified local and global rearrangements of nucleosome occupancy that revealed important roles of nucleosome positioning in cell differentiation.

RESULTS

Nucleosome occupancy maps of ESCs, NPCs and MEFs

We mapped nucleosome positions by genome-wide paired-end sequencing of nucleosomal DNA from mouse ESCs, NPCs and MEFs after digesting the linker DNA between nucleosomes with micrococcal nuclease (MNase) (Online Methods, **Supplementary Fig. 1a–c**, **Supplementary Note**). Examples of the resulting nucleosome coverage maps are depicted in **Figure 1**. We calculated these patterns around transcription factor–binding sites that were determined previously by chromatin immunoprecipitation and DNA sequencing (ChIP-seq) in mouse ESCs[16] or, in the case of the CCCTC-binding factor (CTCF), in both ESCs and MEFs[17,18]. Nucleosome positioning and transcription-factor binding followed a complex relation (**Fig. 1a**). Some transcription-factor sites identified by ChIP-seq in ESCs were nucleosome depleted, with a nucleosome occupancy reduced to 40–80% of that of the flanking regions in all cell types. In contrast, other transcription factors were preferentially bound to nucleosome-enriched regions or displayed distinct patterns that changed during cell differentiation, as described in further detail below. DNA sequence–dependent binding affinities of the histone octamer also contributed to nucleosome positioning. This is inferred from an exemplary comparison of the experimental nucleosome occupancies to those predicted from the DNA sequences[19], in which the computed peaks of nucleosome occupancy correlated well with our experimental nucleosome positions in MEFs but displayed large differences relative to the ESC data set (**Fig. 1b**). Thus, for the latter cell type the intrinsic binding preferences were overwritten by other factors, for example by the presence of CTCF that binds in ESCs to the region shown and is flanked by two well-positioned nucleosomes.

We found many promoter regions to be nucleosome depleted at the transcription start site (TSS), as shown for the *Smarca4* promoter in **Figure 1c**. At this locus, one nucleosome was constitutively absent downstream of the TSS in all three cell lines. An additional nucleosome was also removed upstream of the TSS in ESCs, where this gene was upregulated by about two-fold in comparison to MEFs.

[1]Research Group Genome Organization and Function, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany. [2]BioQuant, Heidelberg, Germany. [3]Division of Theoretical Systems Biology, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany. Correspondence should be addressed to V.B.T. (v.teif@dkfz.de) or K.R. (karsten.rippe@dkfz.de).

Received 10 April; accepted 19 September; published online 21 October 2012; doi:10.1038/nsmb.2419

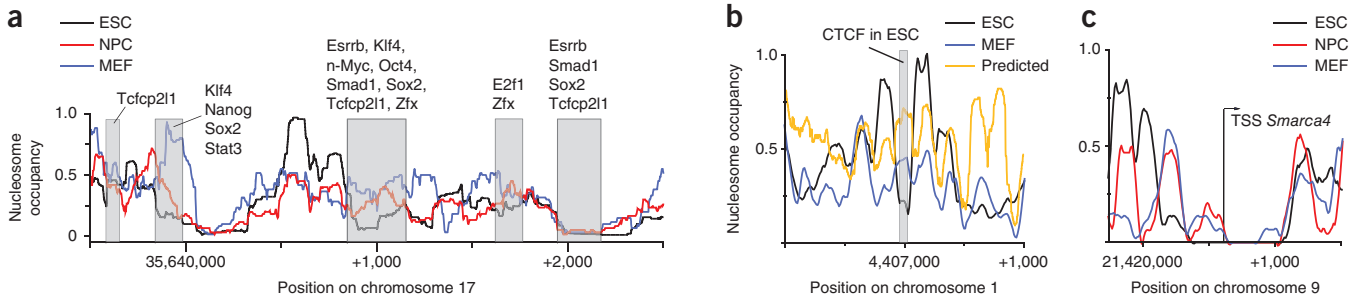


Figure 1 Nucleosome occupancies of exemplary genomic regions in ESCs, NPCs and MEFs. (**a**) Nucleosome occupancies in ESCs (black), NPCs (red) and MEFs (blue) around a putative enhancer region residing in a gene-reach locus, with the nearest gene being *Pou5f1*, which encodes the Oct4 transcription factor. The region shown contains 11 different developmental transcription factors bound in ESCs at the indicated sites, as determined in ref. 16. (**b**) Nucleosome occupancies in ESCs (black) and MEFs (blue) for an exemplary genomic region containing CTCF bound in ESCs, as identified in ref. 18. The experimentally determined nucleosome occupancy is compared to that predicted from the DNA sequence (yellow) according to a previously described algorithm[19]. (**c**) Promoter of the *Smarca4* gene that encodes the chromatin remodeler Brg1, which is involved in development. Color codes for ESC, MEF and NPC are the same as in **a**.

As described in the following, specific features of individual nucleosome occupancy profiles at functional genomic elements like transcription factor–binding sites or TSSs can be evaluated in a genome-wide analysis of the experimental nucleosome occupancy profiles.

Nucleosome occupancy at transcription factor–binding sites

We calculated average nucleosome occupancy profiles for binding sites that were experimentally determined by ChIP-seq in ESCs and that comprised 12 developmentally important transcription factors[16], p300 histone acetyltransferase[18], chromatin remodelers Chd7 (ref. 20) and Brg1 (ref. 21), as well as DNase I–hypersensitivity sites[20]. Four types of patterns were observed (**Fig. 2**). (i) Some transcription factors, such as c-Myc, n-Myc, Zfx and Klf4, were preferentially bound in nucleosome-depleted regions in ESCs (**Fig. 2a**). These regions retained a largely reduced nucleosome occupancy in NPCs and MEFs. A similar pattern was observed for the DNase I–hypersensitivity sites, supporting the previous conclusion that these reflect nucleosome-depleted regions[20]. (ii) For another class of proteins, such as Stat3 and p300, the binding sites resided in nucleosome-depleted regions in ESCs but became preferentially

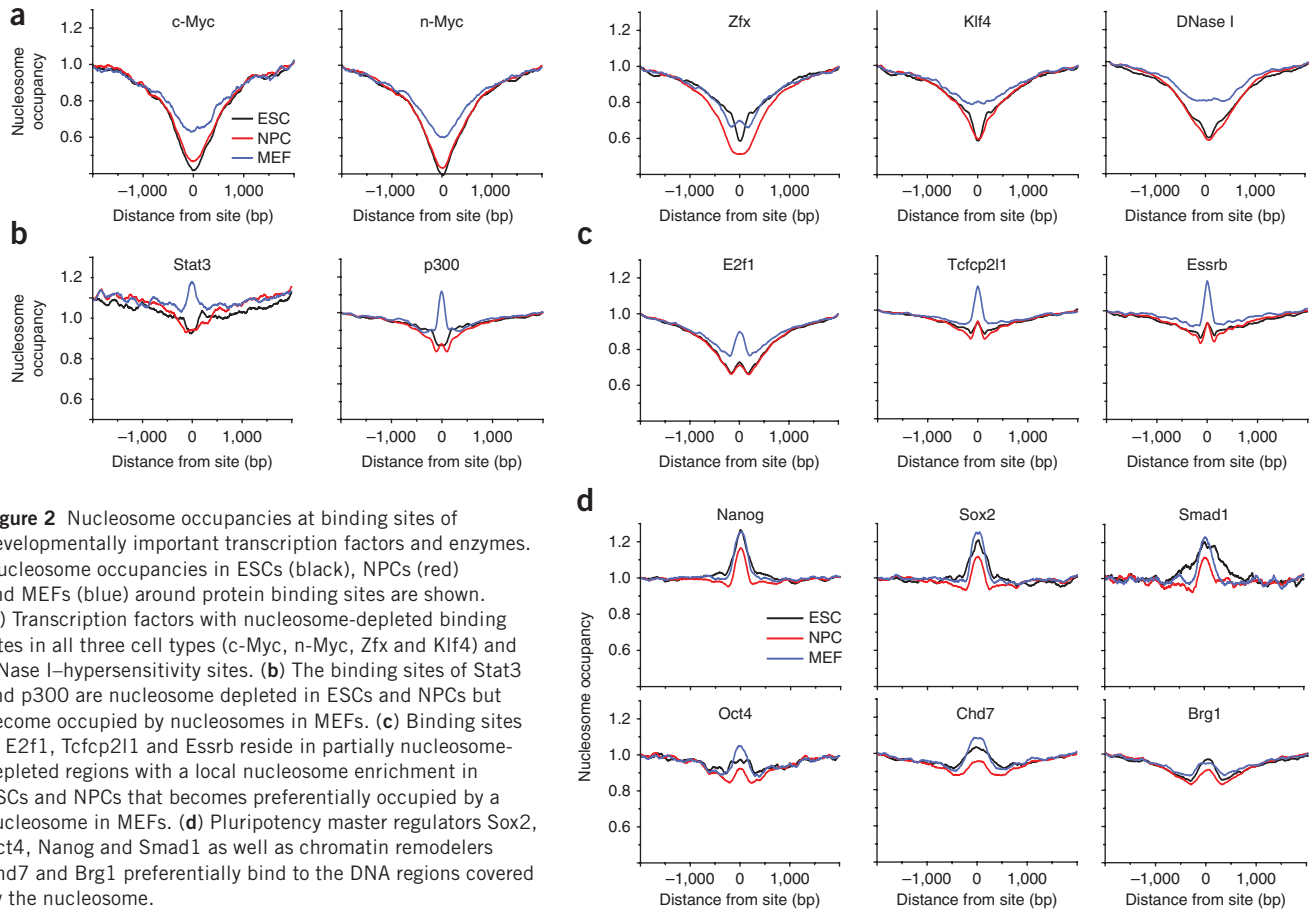


Figure 2 Nucleosome occupancies at binding sites of developmentally important transcription factors and enzymes. Nucleosome occupancies in ESCs (black), NPCs (red) and MEFs (blue) around protein binding sites are shown. (**a**) Transcription factors with nucleosome-depleted binding sites in all three cell types (c-Myc, n-Myc, Zfx and Klf4) and DNase I–hypersensitivity sites. (**b**) The binding sites of Stat3 and p300 are nucleosome depleted in ESCs and NPCs but become occupied by nucleosomes in MEFs. (**c**) Binding sites of E2f1, Tcfcp2l1 and Essrb reside in partially nucleosome-depleted regions with a local nucleosome enrichment in ESCs and NPCs that becomes preferentially occupied by a nucleosome in MEFs. (**d**) Pluripotency master regulators Sox2, Oct4, Nanog and Smad1 as well as chromatin remodelers Chd7 and Brg1 preferentially bind to the DNA regions covered by the nucleosome.

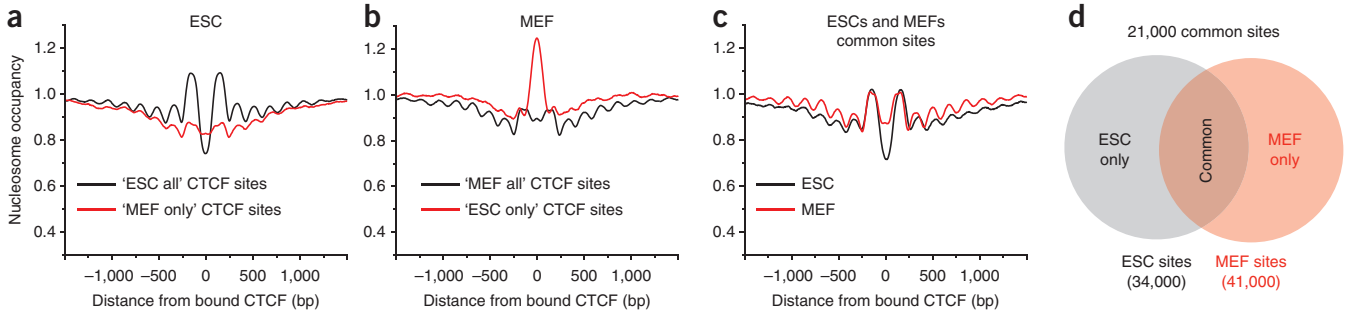


Figure 3 Genome-wide average nucleosome patterns at CTCF binding sites. (**a**–**c**) The experimentally determined CTCF binding sites from ENCODE[18] were split into five groups: (i) CTCF binding sites in ESCs ('ESC all'), (ii) CTCF binding sites in MEFs ('MEF all'), (iii) common CTCF binding sites found simultaneously in both cell lines ('Common'), (iv) CTCF sites detected in ESCs but not found in MEFs ('ESC only') and (v) CTCF sites found in MEFs but not found in ESCs ('MEF only'). Shown are nucleosome occupancy in ESCs at 'ESC all' (black) and at 'MEF only' sites (red) (**a**); in MEFs at sites where CTCF was bound in MEFs (black) or only in ESCs (red) (**b**); and in ESCs (black) and MEFs (red) at the common sites where CTCF was bound simultaneously in both MEFs and ESCs (**c**). (**d**) Venn diagram showing the distribution of CTCF sites between ESCs and MEFs.

occupied by nucleosomes in differentiated cells, possibly because these factors were no longer bound (**Fig. 2b**). (iii) Some proteins, such as E2f1, Tcfcp2l1 and Essrb, showed a more complex pattern with a small nucleosome occupancy peak at the binding sites, surrounded by wider regions of reduced nucleosome occupancy (**Fig. 2c**). This could reflect a regulatory role of nucleosome positioning for those transcription factors that remain expressed at different developmental stages[22]. In addition, binding for this transcription-factor group might require active translocation or eviction of a nucleosome by chromatin remodelers[23,24]. (iv) Another transcription-factor class including the master regulators Nanog, Sox2 and Oct4 had binding sites in ESCs that coincided with well-positioned nucleosomes (**Fig. 2d**). We conclude that these factors can efficiently bind while the DNA target site interacts with a histone octamer, as has been postulated for so-called 'pioneering factors' that initiate cellular programs.

Cell type–dependent CTCF-directed nucleosome positioning

To further explore nucleosome rearrangement around transcription factor–binding sites, we analyzed CTCF binding sites that have been previously mapped in both ESCs and MEFs[17,18] (**Fig. 3a–c**, **Supplementary Fig. 2**). CTCF establishes insulatory or boundary elements to demarcate repressive and active chromatin regions[25] by setting a local boundary organizing 10–20 nucleosomes[26] as well as by bridging distant chromatin regions[27]. CTCF binding sites were recently identified by ChIP-seq by another study[17] and by the ENCODE project[18]. According to the ENCODE data set used in our analysis, the total number of CTCF binding sites is 34,000 in ESCs and 41,000 in MEFs, with only ~30% of sites coinciding in the two cell types (**Fig. 3d**).

Both ESCs and MEFs displayed a nucleosome-depleted region of ~200 base pairs (bp) at the center of the CTCF binding sites, which was somewhat more pronounced in ESCs (**Fig. 3a**). Two nucleosomes were positioned directly adjacent to this site and flanked by up to nine regularly spaced nucleosomes, similar to the pattern reported previously for human T cells[26]. Notably, the subset of CTCF sites that were occupied only in MEFs but not in ESCs was found to already be partly nucleosome depleted in ESCs (**Fig. 3a**). In MEFs, nucleosomes were substantially rearranged around the CTCF sites (**Fig. 3b**). The previous positions of CTCF that were unique to ESCs but not associated with CTCF in MEFs displayed an ~20% increase in nucleosome occupancy. The sites of bound CTCF in MEFs fell into two classes with respect to their nucleosome occupancy (**Fig. 3** and **Supplementary Fig. 2a**). (i) The constitutive CTCF sites present both in ESCs and MEFs were also nucleosome depleted in MEFs but to a lesser extent than in ESCs. This might be related to an ~27% reduction of the ratio of CTCF to core histone H4 expression (**Supplementary Table 1**). Notably, the CTCF sites that were unique to MEFs displayed an additional local peak of nucleosome occupancy (**Supplementary Fig. 2b**). This suggests that CTCF is able to bind to nucleosomal DNA at these sites, possibly through additional interaction partners. In support of this view, an example with CTCF bound in the middle of the nucleosome peak is shown in **Supplementary Figure 2b**. (ii) In addition, a subset of CTCF sites located within enhancer elements displayed a very different pattern (**Supplementary Fig. 2c**).

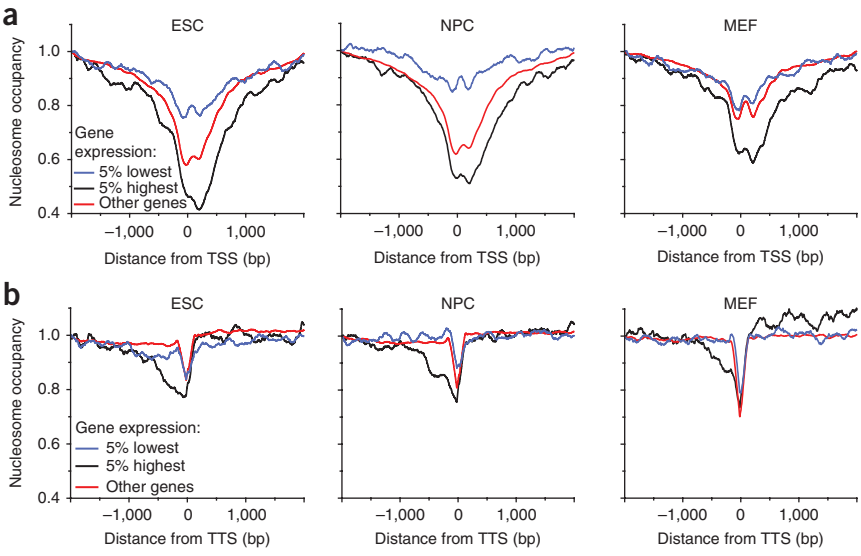


Figure 4 Average nucleosome occupancy patterns around TSSs and TTSs in ESCs, NPCs and MEFs. (**a**) Nucleosome occupancies at the TSS for the 5% highest- (black) and 5% lowest-expressed genes (blue) in comparison to the remaining 90% (red). (**b**) Same as in **a**, but for TTSs.

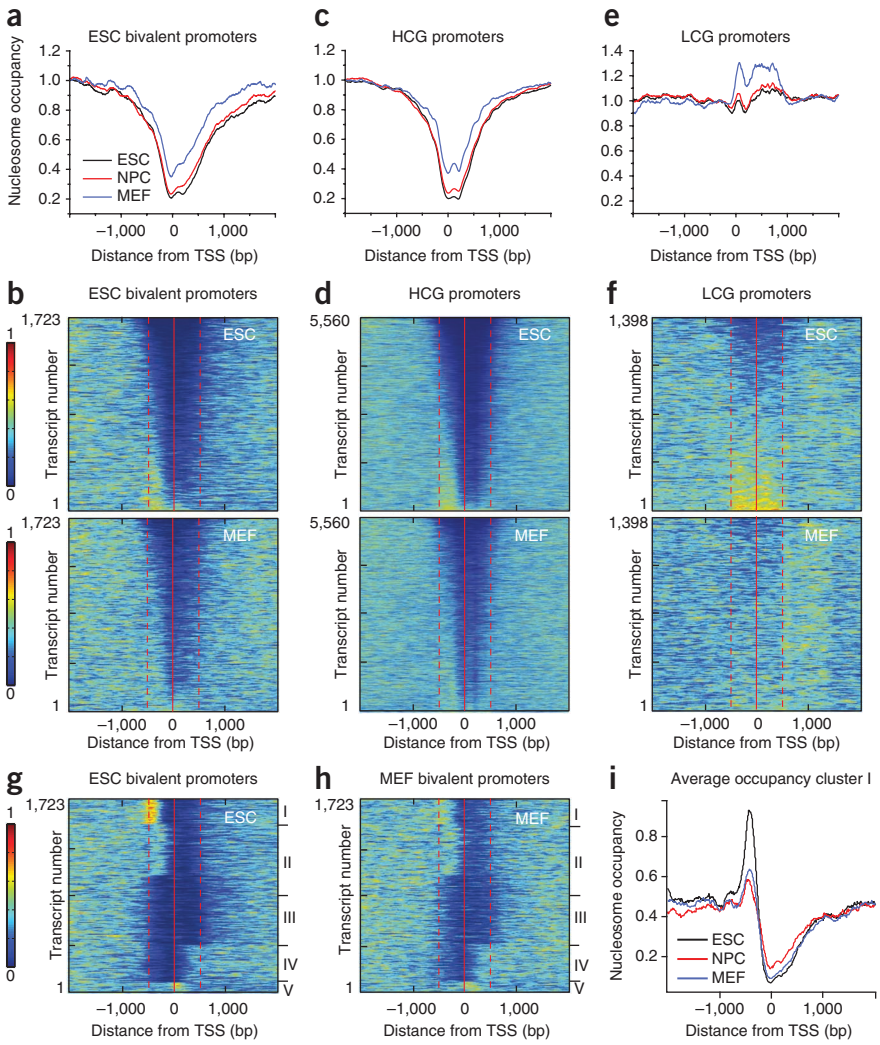


Figure 5 Nucleosome occupancy patterns around TSSs in different classes of promoters. Bivalent promoters that carry both the H3K4me3 and H3K27me3 histone-modification marks in ESCs, HCG promoters and LCG promoters were analyzed in terms of their nucleosome occupancy. Black, ESCs; red, NPCs; blue, MEFs. (**a**) Average nucleosome profiles around bivalent ESC promoters. (**b**) Cluster plots of nucleosome occupancies at the TSS for ESC bivalent promoters in ESCs (top) and MEFs (bottom). Each horizontal line corresponds to the promoter associated with a given transcript. The colors indicate nucleosome occupancy from low (dark blue) to high (red). For ESCs, the sorting was according to the similarity of the nucleosome occupancy profile, starting with the lowest occupancy on top. This order was maintained in MEFs to visualize any nucleosome occupancy changes between ESCs and MEFs. (**c**) Average nucleosome profiles around HCG promoters. (**d**) Cluster plots as in **b**, conducted for HCG promoters. (**e**) Average nucleosome profiles around LCG promoters. (**f**) Cluster plots as in **b**, conducted for LCG promoters. (**g**) Hierarchical cluster plot of bivalent promoters in ESCs split into five groups. Cluster I has a nucleosome covering the region from about –500 to –350 bp relative to the TSS. (**h**) Hierarchical cluster plot for MEFs with the same promoters and ordering as in **g**. It is apparent that the –500 to –350 bp region in cluster I looses a nucleosome. (**i**) Average nucleosome occupancy of the genes in cluster I in ESCs, NPCs and MEFs.

Nucleosome occupancies at TSSs and transcription end sites

To characterize the nucleosome occupancy at promoters, we aligned the maps of mouse transcripts at their TSSs and clustered them according to their expression levels for each cell type into active (top 5% of expression level), inactive (bottom 5% of expression level) and the remaining 90% of the transcripts. We observed a broad nucleosome-depleted region downstream of the TSS for ESCs, NPCs and MEFs (**Fig. 4a**). It extended into the gene body similarly to the pattern reported for mouse hepatocytes[15] but very differently from the typical patterns observed in yeast and invertebrates[5–7,28,29]. The nucleosome occupancy profiles in ESCs, NPCs and MEFs were dependent on the gene expression level. For the most active genes, the nucleosome-depleted region became wider and deeper. **Supplementary Figure 3** shows the occupancy maps of the 5% of genes with the highest and lowest expression levels. In general, the TSS nucleosome occupancy displayed a significant anticorrelation with expression of the corresponding transcripts (**Supplementary Table 2**), indicating that a reduced nucleosome occupancy at the promoter favors gene expression.

Next, we grouped average nucleosome profiles at the transcription termination sites (TTSs) of the 5% highest, the 5% lowest and the remaining 90% of genes in ESCs, NPCs and MEFs, according to gene expression level (**Fig. 4b**). Inactive genes were characterized by a relatively small nucleosome-depleted region around the TTS, possibly due to a nucleosome-excluding DNA sequence[30]. This nucleosome-depleted region was largely increased upstream of the TTS for the highly active genes in all cell types (**Fig. 4b** and **Supplementary Fig. 4**) and was different from the nucleosome-depleted region downstream of the TTS reported previously for yeast[31–33]. Thus, both TSSs and TTSs have unique nucleosome signatures that might reflect more complex regulatory mechanisms than those found in simpler eukaryotes.

Promoter nucleosome occupancy and gene expression changes

To investigate whether changes in nucleosome occupancy between ESCs and MEFs were correlated with gene expression changes, we evaluated and identified three different promoter classes (**Fig. 5**, **Supplementary Fig. 5** and **Supplementary Tables 3** and **4**). The first class was defined by the simultaneous presence of the trimethylation modification of histone H3 at lysine residue 4 (H3K4me3) and at lysine 27 (H3K27me3) in ESCs and is referred to as 'bivalent' promoters[34,35] (**Fig. 5a**,**b** and **Supplementary Fig. 5**). Two other classes were distinguished according to their DNA sequence composition as high CpG (HCG) (**Fig. 5c**,**d**) or low CpG (LCG) promoters (**Fig. 5e**,**f**)[36]. For bivalent and HCG promoters, the nucleosome occupancy profiles were characterized by a strong nucleosome-depleted region (**Fig. 5a**–**d**) similar to the average TSS patterns (**Fig. 4a**) and with no substantial differences between the three cell types. In contrast, the nucleosome occupancy around the TSS was high for LCG promoters (**Fig. 5e**,**f**) and for promoters that carried only the H3K27me3 modification (**Supplementary Fig. 5**). We then sorted the TSSs within each class according to their average nucleosome occupancy

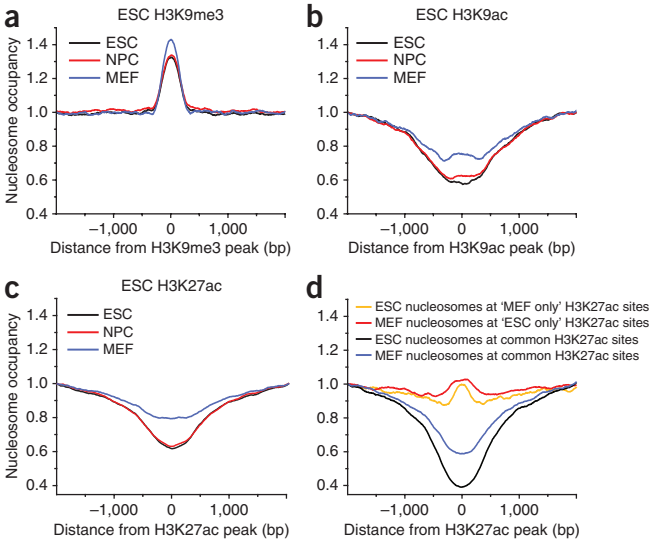


Figure 6 Average nucleosome occupancy patterns around clusters of H3K9me3, H3K9ac and H3K27ac histone modifications. Nucleosome occupancies in ESCs (black), NPCs (red) and MEFs (blue) aligned with respect to the cluster centers of the indicated histone modifications. (**a**–**c**) Nucleosome occupancy at the clusters of H3K9me3 (**a**), H3K9ac (**b**) and H3K27ac (**c**) found in ESCs. (**d**) Comparison of nucleosome occupancy profiles around the H3K27ac modification in ESCs and MEFs. H3K27ac clusters were split into those that were present only in ESCs ('ESC only') or only in MEFs ('MEF only') and 'common sites' that were present simultaneously in ESCs and MEFs.

in ESCs in the region at −500 to 500 bp around the TSS and compared these to MEFs while keeping the same ordering. The overall pattern remained very similar between ESCs and MEFs for bivalent (**Fig. 5b**) and HCG promoters (**Fig. 5d**) but changed for a large number of LCG promoters (**Fig. 5f**).

To test whether nucleosome occupancy changes in the region at −500 to 500 bp were linked to gene expression changes between ESCs and MEFs, we conducted a correlation analysis of the corresponding $\log_2$ ratios (**Supplementary Tables 3** and **4**). In this analysis we found no simple correlation between the two parameters except for LCG promoters. In addition, for the two small groups of ESC bivalent promoters that had transcripts detected by RNA-seq in ESCs but not in MEFs (36 genes) and ESC H3K27me3 promoters that were found to be expressed in MEFs but not in ESCs (20 genes), the data were indicative of an increase of nucleosome occupancy at the TSS during silencing.

We further dissected the relation between nucleosome occupancy and gene expression changes for subgroups of bivalent (H3K4me3 and H3K27me3) promoters in ESCs. In NPCs these resolve into promoters that either carry only the H3K4me3 or only the H3K27me3 mark[34,35]. For the averaged profiles, we did not observe substantial changes of nucleosome occupancy (**Supplementary Fig. 5b**). Clustering of bivalent promoters according to their nucleosome occupancy pattern revealed one group, designated as cluster I, which showed a well-positioned nucleosome occupancy peak in the region from −500 to −350 bp upstream of the TSS (**Fig. 5g**). This nucleosome was preferentially removed in both NPCs and MEFs (**Fig. 5h**,**i**) with a concomitant average gene expression increase of 2.2-fold in NPCs and 4.6-fold in MEFs, similar to the average gene expression change observed for all bivalent promoters. Thus, activation of cluster I genes could involve the complete removal of a nucleosome instead of changing its associated histone modifications. Within this cluster, gene ontology categories were enriched that are associated with differentiated cell function (blood-vessel development, 9 genes, $P < 4.7 \times 10^{-3}$; positive regulation of transcription and gene expression, 11 genes, $P < 4.1 \times 10^{-2}$; cell migration and cell motility, 7 genes, $P < 7.8 \times 10^{-2}$; axonogenesis, neuron projection morphogenesis and cell morphogenesis, 6 genes, $P < 7.9 \times 10^{-2}$; **Supplementary Table 5**).

Histone modification–dependent nucleosome occupancies

Next, we investigated whether nucleosome occupancies changed between transcriptionally active or inactive chromatin regions. Histone modifications were determined by ChIP-seq, and well-defined peaks ($P < 10^{-5}$) were selected. We identified about 10,000 clusters for each histone mark studied. These included the bona fide repressive trimethylation modification of histone H3 at lysine residue 9 (H3K9me3), as well as the permissive acetylation of histone H3 at either lysine residue 9 (H3K9ac) or at lysine 27 (H3K27ac) in ESCs and MEFs. Average nucleosome occupancy patterns in ESCs, NPCs and MEFs were calculated around the centers of these clusters (**Fig. 6**). We found that H3K9me3 clusters displayed increased nucleosome occupancy (**Fig. 6a**), whereas H3K9ac and H3K27ac clusters showed the opposite trend (**Fig. 6b**,**c**). H3K27ac sites common to ESCs and MEFs were substantially depleted of nucleosomes in both ESCs and MEFs (**Fig. 6d**). However, genomic positions that were acetylated in ESCs but not in MEFs had a pronounced peak of nucleosome occupancy in MEFs. The same was true for the MEF-specific H3K27ac

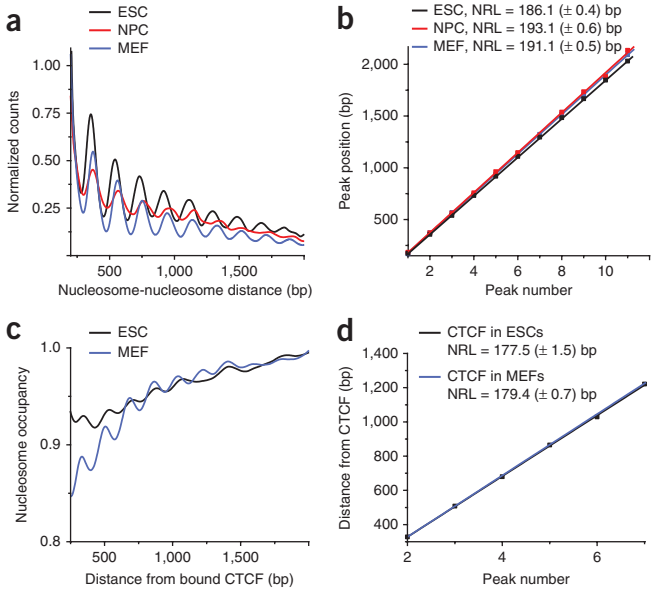


Figure 7 Nucleosome repeat length in ESCs, NPCs and MEFs. (**a**) Histogram of distances between all nucleosome start positions. Peaks reflect the distances between nearest-neighbor nucleosomes, next-nearest neighbors, and so on. The largest frequency peak found for small distance values around zero is not relevant for the NRL analysis because it reflects the experimental uncertainty for determining the boundary of the same nucleosome. (**b**) Plot of the peak position along the *x* axis versus the peak number. The slope of the line fitted through the plot yields the NRL for a given cell type. (**c**) Determination of the local NRL for the region around CTCF binding sites in ESCs and MEFs. The centrally positioned nucleosomes flanking the CTCF site were excluded from the analysis, and the adjacent 8 nucleosomes on the right side were evaluated. (**d**) Linear fit for NRL determination in regions of bound CTCF in ESCs and MEFs for the peaks identified in **c**.

sites: these positions had increased nucleosome occupancy in ESCs, where they were not acetylated. Thus, histone modifications could mark nucleosomes for changes in their density along the DNA in the corresponding regions.

Increase of nucleosome repeat length during differentiation

An essential parameter that describes the primary chromatin organization is the nucleosome repeat length (NRL), which is the average distance between two neighboring nucleosomes. We determined NRLs according to a previously described method[9]. It is based on calculating the frequency of nucleosome distances between the starts of all mononucleosomal DNA fragments and then analyzing the preferred distances between the nearest-neighbor nucleosomes, next-nearest neighbors, etc. The resulting plots yielded well-defined peaks for the preferred internucleosome distances (**Fig. 7a**). From the plot of peak position and corresponding nucleosome number, we obtained values of 186.1 ± 0.4 bp (ESCs), 193.1 ± 0.6 bp (NPCs) and 191.1 ± 0.5 bp (MEFs): that is, the average NRL increased by 5–7 bp during differentiation (**Fig. 7b**). On the basis of previous findings, the changes in NRL could involve a change in the molar ratio of linker to core histones[37]. We found that for only the linker histone variants H1.0 and H1.7, gene expression in relation to that of the core histones was raised substantially in both NPCs and MEFs (**Supplementary Table 1**). Accordingly, these variants might be particularly important for inducing an NRL increase during differentiation.

Nucleosome position distances at specific genomic loci displayed large local variations from these average NRLs, for example at the TSS and TTS, due to the binding of other protein factors. Moreover, in a 4-kb region of orderly packed nucleosomes around CTCF binding sites, the NRL was reduced to 177.5 ± 1.5 bp for ESCs and 179.4 ± 0.7 bp for MEFs: that is, both values were ~10 bp smaller than the corresponding genome-wide NRLs (**Fig. 7c**,**d**).

DISCUSSION

Our analysis of nucleosome positioning in mouse ESCs in comparison to their lineage-committed NPC and MEF counterparts revealed distinct profiles at functional genomic elements that are relevant for cell differentiation. The analysis of nucleosome occupancy at transcription factor–binding sites indicated the presence of different chromatin interaction mechanisms (**Figs. 1**, **2** and **3**); some binding sites were constitutively depleted of nucleosomes in all three cell types (**Fig. 2a**). This might be an important feature of a certain set of constitutive sites that are always competent for transcription-factor binding if the appropriate factor becomes expressed. In contrast, for other transcription factors the nucleosome occupancy at target sites showed little correlation with transcription-factor binding, which suggested that these transcription factors can bind to nucleosomal DNA or that binding occurs in only a small fraction of the cells at any given point in time (**Fig. 2d**). In a third group, transcription-factor and histone-octamer binding appeared to be in a competitive equilibrium, in which an increase of the transcription-factor concentration during development could be sufficient to displace the nucleosome from a position that interferes with binding to result in occupancy profiles like those observed in **Figure 2b**,**c**. In addition, chromatin-remodeling complexes like Chd7 and Brg1 (**Fig. 2d**) could bind to a nucleosome and actively translocate it along the DNA to a new position, as discussed previously[23]. Because the DNA occupancy of developmental transcription factors is highly predictive of gene expression in mouse ESCs[38], any modulation of this parameter by nucleosome occupancy at a given transcription factor–binding site could directly influence the expression of the target genes.

A particularly interesting example of the complex relationship between protein binding and nucleosome positioning was revealed here for CTCF (**Fig. 3** and **Supplementary Fig. 2**). Consistent with the previous findings[39], the CTCF binding sites in ESCs were located in regions with reduced nucleosome occupancy and acted as a nucleosome boundary element to position adjacent nucleosomes. Notably, some CTCF binding sites specific for MEFs appear to be predisposed in ESCs for later CTCF binding (**Fig. 3a**). In contrast, CTCF binding sites unique to ESCs became occupied with nucleosomes in MEFs, which would be consistent with the dissociation of CTCF and possibly related to its decreased expression in MEFs (**Supplementary Table 1**). As reported previously, CTCF binding sites in general have an intrinsically high affinity for the histone octamer, which could promote their incorporation into a nucleosome if CTCF dissociates[9]. Unexpectedly, a fraction of CTCF proteins in MEFs was apparently associated with nucleosomes (**Supplementary Fig. 2a**,**b**), and for CTCF binding sites located within enhancers, no regular nucleosome positioning pattern was detected (**Supplementary Fig. 2c**). Thus, our results suggest that multiple modes of CTCF interaction with chromatin exist, which might involve other protein factors or RNAs that mediate CTCF binding to nucleosomal DNA at certain sites during differentiation.

In order to investigate the gene-specific functions of nucleosome positioning, we conducted an analysis of nucleosome occupancy at the TSS (**Figs. 4a** and **5**). The profile of the nucleosome-depleted region varied for different classes of promoters that were selected either on the basis of the presence of the H3K4me3 and H3K27me3 histone modifications or the CpG content of the DNA sequence (**Fig. 5** and **Supplementary Fig. 5**). For the majority of promoters, a nucleosome-depleted region centered around +100 bp was present that became more pronounced for highly active genes. This profile might be related to the presence of RNA polymerase II that has been mapped recently at the promoters of mouse ESCs[40]. RNA polymerase II could be present in either a transcriptionally engaged form or bound in a stalled state that requires additional factors for initiation of transcription[36]. Notably, the transcript ends displayed a decrease of nucleosome occupancy toward the gene body for the most active genes (**Fig. 4b**).

The average TSS nucleosome pattern determined here was similar to nucleosome profiles reported previously for mouse hepatocytes[15] and selected human promoters[9,10,41] that were also characterized by a rather broad nucleosome-depleted region. This pattern is markedly different from those previously reported for simpler eukaryotes, in which a single nucleosome was missing upstream of the TSS and downstream of the TTS, and this was followed by an oscillatory pattern of several regularly positioned nucleosomes[5–7,28,29]. In addition, a nucleosome-depleted region extending into the gene body was found at the TTSs, and it became more pronounced with increased gene expression (**Fig. 4b**). This feature could be related to a coupling of the TTS with 3′ polyadenylation of the transcript[42].

When evaluating all promoters within one cell type, we found that an increased nucleosome occupancy in the region of 500 bp around the TSS was correlated with a reduction of transcription (**Supplementary Table 2**). Although no corresponding anticorrelated changes of promoter nucleosome occupancy and gene expression were found between ESCs and MEFs for the majority of genes (**Supplementary Tables 3** and **4**), we identified certain specific groups of promoters that displayed such a behavior. This qualifies them as potential candidates for a regulatory mechanism that would operate through nucleosome repositioning during differentiation. For example, the bivalent promoters in ESCs contained a group of promoters with a specific nucleosome occupancy profile. This cluster showed a correlation between the loss

of a nucleosome, at position −500 to −350 bp around the TSS in both NPCs and MEFs, accompanied by an increase in gene expression (**Fig. 5g–i**). In addition, other ESC bivalent and H3K27me3-only promoters as well as LCG promoters displayed anticorrelated relations between nucleosome occupancy and gene expression changes during differentiation (**Supplementary Tables 3** and **4**).

These changes might be directly related to the addition or removal of certain histone marks, as concluded from our analysis of exemplary histone modifications with respect to nucleosome occupancy (**Fig. 6**). H3K9me3 clusters in ESCs were found to be nucleosome enriched, whereas H3K9ac and H3K27ac clusters were nucleosome depleted. H3K9me3 and H3K9ac are particularly notable because H3K9me3 is highly correlated with local mutation rates in cancer cells, whereas H3K9ac is strongly anticorrelated[43]. Together with recent findings that chromatin regions characterized by different histone modifications vary in their nucleosome repeat length[9], our data link histone modifications with important structural functions with respect to nucleosome positioning and occupancy. As reviewed recently, chromatin-remodeling complexes recognize a variety of histone modifications[24]. Thus, it is tempting to speculate that these molecular machines are involved in changes of nucleosome occupancy at transcription factor–binding sites and promoters after marking a given nucleosome with specific histone modification signals.

Finally, we observed genome-wide changes of the primary chromatin structure during cell differentiation, as reflected in the increase of the average NRL in ESCs of 186.1 ± 0.4 bp by 7 bp (NPCs) and 5 bp (MEFs) (**Fig. 7**). As reported previously, this NRL change is related to an increase in the ratio of linker to core histones[37]. Although mouse ESCs have a histone H1/nucleosome ratio of 0.46 (ref. 37), this parameter increases to 0.75–0.83 in various differentiated mouse tissues[44,45]. The upregulated gene expression of linker histone variants H1.0 and H1.7 in NPCs and MEFs versus ESCs (**Supplementary Table 1**) indicates that these factors might be particularly important for the change in the NRL. It is noted that the 5–7-bp difference in NRL observed between ESCs and NPCs and MEFs could have large effects on the folding properties of the nucleosome chain because the helical phasing of the DNA double helix would relocate neighboring nucleosomes by a torsional angle of about 36° per additional base pair. In agreement with this view, large differences in the chromatin folding properties as a function of NRL have been observed experimentally as, for example, reported in ref. 46. In addition, our calculation for selected short genomic regions revealed that the NRL shows local variation, as for example in the 4-kb region surrounding CTCF binding sites that had an ~10 bp smaller NRL than the genome-wide average value.

In summary, we identified a number of substantial rearrangements of nucleosome positions at different functional genomic elements like transcription factor–binding sites and promoters that are likely to modulate protein binding to these regions. In addition, global changes of nucleosome density that occur throughout the genome during cell lineage commitment of mouse ESCs could also affect the DNA accessibility by changing the folding of the nucleosome chain. Accordingly, we conclude that the cell type–specific organization of nucleosomes on identical genomes represents an additional regulatory layer that controls DNA access of protein factors for selecting tissue-specific gene expression programs.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. MNase-seq, ChIP-seq and RNA-seq data have been deposited to the GEO database under the accession number GSE40896.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We are grateful to A. Valouev and R. Chereji for help with the algorithms for calculations of NRL and average TSS patterns, respectively; to M. Gerstein for advice on data processing; to G. Längst, G. Wedemann and K. Fejes Tóth for discussions; and to the Deutsches Krebsforschungszentrum Sequencing Core Facility for conducting the sequencing. This work was funded within project EpiGenSys by the German Federal Ministry of Education and Research (BMBF) as a partner of the ERASysBio+ initiative in the EU FP7 ERA-NET Plus program through grant 0315712A to K.R. Computational resources and data storage were provided by grants from the BMBF (01IG07015G, Services@MediGRID) and the German Research Foundation (DFG INST 295/27-1). V.B.T. acknowledges the support from the Heidelberg Center for Modeling and Simulation in the Biosciences and a Deutsches Krebsforschungszentrum intramural grant, and Y.V. was supported by BMBF MedSys grant 0315409E to T.H.

AUTHOR CONTRIBUTIONS

V.B.T. and K.R. designed the research. M.C.-H., J.-P.M. and C.M. performed experiments. V.B.T., Y.V., J.-P.M., T.H. and K.R. analyzed data. V.B.T., T.H. and K.R. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nsmb.2419.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Sadeh, R. & Allis, C.D. Genome-wide "re"-modeling of nucleosome positions. *Cell* **147**, 263–266 (2011).
2. Rando, O.J. & Winston, F. Chromatin and transcription in yeast. *Genetics* **190**, 351–387 (2012).
3. Zhang, Z. & Pugh, B.F. High-resolution genome-wide mapping of the primary structure of chromatin. *Cell* **144**, 175–186 (2011).
4. Cui, K. & Zhao, K. Genome-wide approaches to determining nucleosome occupancy in metazoans using MNase-Seq. *Methods Mol. Biol.* **833**, 413–419 (2012).
5. Yuan, G.C. *et al.* Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309**, 626–630 (2005).
6. Ioshikhes, I.P., Albert, I., Zanton, S.J. & Pugh, B.F. Nucleosome positions predicted through comparative genomics. *Nat. Genet.* **38**, 1210–1215 (2006).
7. Segal, E. *et al.* A genomic code for nucleosome positioning. *Nature* **442**, 772–778 (2006).
8. Hu, G. *et al.* Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1. *Genome Res.* **21**, 1650–1658 (2011).
9. Valouev, A. *et al.* Determinants of nucleosome organization in primary human cells. *Nature* **474**, 516–520 (2011).
10. Ott, C.J. *et al.* Nucleosome occupancy reveals regulatory elements of the CFTR promoter. *Nucleic Acids Res.* **40**, 625–637 (2012).
11. Schones, D.E. *et al.* Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887–898 (2008).
12. Zhang, L., Ma, H. & Pugh, B.F. Stable and dynamic nucleosome states during a meiotic developmental process. *Genome Res.* **21**, 875–884 (2011).
13. Tirosh, I., Sigal, N. & Barkai, N. Widespread remodeling of mid-coding sequence nucleosomes by Isw1. *Genome Biol.* **11**, R49 (2010).
14. Moshkin, Y.M. *et al.* Remodelers organize cellular chromatin by counteracting intrinsic histone-DNA sequence preferences in a class-specific manner. *Mol. Cell Biol.* **32**, 675–688 (2012).
15. Li, Z., Schug, J., Tuteja, G., White, P. & Kaestner, K.H. The nucleosome map of the mammalian liver. *Nat. Struct. Mol. Biol.* **18**, 742–746 (2011).
16. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).
17. Martin, D. *et al.* Genome-wide CTCF distribution in vertebrates defines equivalent sites that aid the identification of disease-associated genes. *Nat. Struct. Mol. Biol.* **18**, 708–714 (2011).
18. Shen, Y. *et al.* A map of the *cis*-regulatory sequences in the mouse genome. *Nature* **488**, 116–120 (2012).
19. Kaplan, N. *et al.* The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**, 362–366 (2009).
20. Schnetz, M.P. *et al.* CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression. *PLoS Genet.* **6**, e1001023 (2010).
21. Ho, L. *et al.* An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proc. Natl. Acad. Sci. USA* **106**, 5187–5191 (2009).

22. Zielke, N. *et al.* Control of *Drosophila* endocycles by E2F and CRL4(CDT2). *Nature* **480**, 123–127 (2011).
23. Teif, V.B. & Rippe, K. Predicting nucleosome positions on the DNA: combining intrinsic sequence preferences and remodeler activities. *Nucleic Acids Res.* **37**, 5641–5655 (2009).
24. Erdel, F., Krug, J., Langst, G. & Rippe, K. Targeting chromatin remodelers: signals and search mechanisms. *Biochim. Biophys. Acta* **1809**, 497–508 (2011).
25. Cuddapah, S. *et al.* Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* **19**, 24–32 (2009).
26. Fu, Y., Sinha, M., Peterson, C.L. & Weng, Z. The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet.* **4**, e1000138 (2008).
27. Handoko, L. *et al.* CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat. Genet.* **43**, 630–638 (2011).
28. Mavrich, T.N. *et al.* Nucleosome organization in the *Drosophila* genome. *Nature* **453**, 358–362 (2008).
29. 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).
30. Segal, E. & Widom, J. Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr. Opin. Struct. Biol.* **19**, 65–71 (2009).
31. Zhang, Y. *et al.* Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions *in vivo*. *Nat. Struct. Mol. Biol.* **16**, 847–852 (2009).
32. Weiner, A., Hughes, A., Yassour, M., Rando, O.J. & Friedman, N. High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. *Genome Res.* **20**, 90–100 (2010).
33. Fan, X. *et al.* Nucleosome depletion at yeast terminators is not intrinsic and can occur by a transcriptional mechanism linked to 3′-end formation. *Proc. Natl. Acad. Sci. USA* **107**, 17945–17950 (2010).
34. Ku, M. *et al.* Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* **4**, e1000242 (2008).
35. Mikkelsen, T.S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
36. Lenhard, B., Sandelin, A. & Carninci, P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat. Rev. Genet.* **13**, 233–245 (2012).
37. 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).
38. Cheng, C. & Gerstein, M. Modeling the relative relationship of transcription factor binding and histone modifications to gene expression levels in mouse embryonic stem cells. *Nucleic Acids Res.* **40**, 553–568 (2012).
39. Kanduri, M. *et al.* Multiple nucleosome positioning sites regulate the CTCF-mediated insulator function of the H19 imprinting control region. *Mol. Cell Biol.* **22**, 3339–3344 (2002).
40. Min, I.M. *et al.* Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes Dev.* **25**, 742–754 (2011).
41. Pham, C.D., Sims, H.I., Archer, T.K. & Schnitzler, G.R. Multiple distinct stimuli increase measured nucleosome occupancy around human promoters. *PLoS ONE* **6**, e23490 (2011).
42. Zhao, J., Hyman, L. & Moore, C. Formation of mRNA 3′ ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.* **63**, 405–445 (1999).
43. Schuster-Böckler, B. & Lehner, B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488**, 504–507 (2012).
44. Woodcock, C.L., Skoultchi, A.I. & Fan, Y. Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Res.* **14**, 17–25 (2006).
45. Fan, Y. *et al.* H1 linker histones are essential for mouse development and affect nucleosome spacing *in vivo*. *Mol. Cell Biol.* **23**, 4559–4572 (2003).
46. Correll, S.J., Schubert, M.H. & Grigoryev, S.A. Short nucleosome repeats impose rotational modulations on chromatin fibre folding. *EMBO J.* **31**, 2416–2426 (2012).

ONLINE METHODS

Isolation of nucleosomes. ESCs from 129P2/Ola mice[47] were cultured in ESGRO complete medium (Millipore). Differentiation of ESCs into neuronal precursors was induced by formation of embryoid bodies in embryoid body–formation medium (Millipore) and treatment with 5 µM retinoic acid for 4 d. Neuronal embryoid bodies were dissociated and seeded on Matrigel (BD Biosciences) in neuronal stem cell medium (PAN) for 4 d. MEFs were generated from pregnant 129P2/Ola E13.5 mice and cultured in DMEM supplemented with 10% FCS and glutamine for up to passage 5. For MNase digestion, cells were harvested and resuspended in low-salt buffer (10 mM HEPES, pH 8, 10 mM KCl, 0.5 mM DTT) at 4 °C. After disruption of the cells with a dounce homogenizer, the nuclei were collected by centrifugation and washed once with the MNase Buffer (10 mM Tris-HCl, pH 7.5, 10 mM $CaCl_2$), resuspended in the MNase Buffer and digested with 0.5 units MNase (Fermentas) per microliter and incubated for 6–11 min at 37 °C. The MNase digestion was stopped by putting the samples on ice and adding EDTA to a concentration of 10 mM. After digestion with 0.1 µg µl^{-1} RNase A (Fermentas) and removal of protein by phenol-chloroform extraction, the DNA was ethanol precipitated, and the resulting DNA pellet was dissolved in H_2O. DNA fragments corresponding to mononucleosomes or dinucleosomes were separated on a 2% agarose gel by using an E-Gel electrophoresis system (Life Technologies). The libraries for sequencing were prepared according to the standard protocol for the Illumina HiSeq2000 sequencing platform.

Deep sequencing of nucleosomal DNA. High-throughput paired-end sequencing of at least 50-bp read length was performed on the Illumina HiSeq2000 platform at the DKFZ sequencing core facility in Heidelberg, Germany. We mapped about 150 million nucleosome positions per sequencing reaction and used in the final analysis three biological-replicate experiments for ESCs and two replicate experiments for each of NPCs and MEFs, yielding a total of 300 million–450 million nucleosome positions per cell type. In line with the previous studies[48,49], we observed that the chromosome-wide nucleosome density was dependent on the average GC content, which was anticorrelated with the MNase preferences found with purified genomic mouse DNA (**Supplementary Fig. 1**, **Supplementary Note**). For mapping the position of individual nucleosomes, MNase sequence preferences were found to be negligible, in agreement with a recent study[50]. Following previous findings described in refs. 32 and 51, we checked the dependence of the nucleosome maps on the level of MNase digestion. Using slightly different levels of MNase digestions in four replicate experiments in ESCs, we obtained average mononucleosome fragment lengths around 150 bp, 155 bp, 160 bp and 180 bp, with ~150 million mapped reads in each reaction. The changes of the integral parameters such as the nucleosome repeat length during the cell differentiation were found to be independent of the degree of MNase digestion. However, in line with previous studies[31,48,50], different levels of MNase digestion affected nucleosome distributions, with individual nucleosome peaks sometimes missing or appearing, without clear indications that one of the samples with 150-bp, 155-bp or 160-bp average length was a better representation of the situation *in vivo*. For MNase digestion with 180-bp average fragment length, a large fraction of the linker remained undigested, leading to a largely reduced coverage of individual nucleosome positions. Accordingly, the nucleosome occupancy maps used here were generated from combining only samples with MNase digestions that had an average mononucleosome fragment length between 150–160 bp.

Data analysis of nucleosome occupancies. DNA reads were aligned on the mm9 assembly version of the mouse genome, with Bowtie[52] reporting unique hits with up to two mismatches. The nucleosome occupancy maps were calculated with custom-made Perl scripts by counting how many reads covered a given DNA base pair (**Supplementary Fig. 1b**). Sites with artificially high coverage were considered as artifacts and excluded from the analysis. No further peak calling or smoothing was conducted. In addition, no assumptions on the length of the nucleosomal DNA had to be made to derive the nucleosome occupancy maps, as nucleosome boundaries were determined on both sides of the nucleosome by paired-end sequencing. The nucleosome signatures at transcription factor–binding sites and TSSs or TTSs, respectively, were calculated as the sum of nucleosome occupancies in a window of –2,000 to 2,000 bp around a given site. For each gene, the sum of reads was normalized to 1. Then the averaged nucleosome profile was normalized to yield the nucleosome occupancy equal to 1 at position –2,000 bp[53]. For nucleosome alignment around CTCF binding sites, we analyzed the data set from a previous study[17] downloaded from the GEO archive (GSE27944) and the data sets from ENCODE[18] downloaded from the UCSC Genome Browser (accession codes: wgEncodeEM001703 and wgEncodeEM001698), which resulted in qualitatively similar patterns. Only the patterns obtained with ENCODE data are reported here. For nucleosome alignment around p300 sites, we used the ENCODE data set wgEncodeLicrTfbsEsb-4P300ME0C57bl6StdPk. For nucleosome alignment around binding sites of 12 developmental transcription factors, the analysis was conducted according to the data set from ref. 16, which was initially mapped to the mm8 genome build and converted to mm9 by using the liftOver tool of the UCSC Genome Browser. The histone-modification data from refs. 34 and 35 were also converted from mm8 to mm9 before the analysis. Brg1 ChIP-seq data from ref. 21 (GEO archive GSE14344) were reclustered with MACS[54] using a $P = 10^{-5}$ cutoff for peak detection. Chd7 ChIP-seq data were from ref. 20 (GEO archive GSM558674). DNase I–hypersensitivity raw data from the latter study were provided by the authors and mapped and clustered as described above.

For nucleosome alignment around the TSS and TTS, we used the Eldorado gene annotation provided in the Genomatix Genome Analyzer software (Genomatix)[55]. Alignments with the RefSeq gene annotation resulted in similar patterns. Nucleosome-occupancy cluster plots for visualizing multiple transcripts were generated in Matlab (Mathworks). These profiles were based on the average occupancy at the TSS in the region from position –500 to +500 bp. Hierarchical clustering was done according to the Ward's minimum variance method implemented in Matlab, which computes mutually exclusive groups of occupancy profiles with minimum within-cluster variance[56]. The resulting clusters were analyzed with the DAVID gene annotation clustering tool[57].

The nucleosome repeat length was calculated essentially as described previously[9], with the following modifications: A histogram of the number of occurrences of nucleosome distances from 1–3,000 bp between all nucleosomes was computed and smoothed with a 50-bp window. The NRL was then determined from a linear fit of the detected peak positions versus the nucleosome number. Up to 12 peaks could be identified in this analysis. To make calculations more robust, a threshold of 20 reads was set to remove artificially enriched nucleosome fragments starting at a given genomic position. These nucleosomes were given a statistical weight equal to 20. Less-abundant nucleosome fragments entered the calculation with the weights equal to the number of their occurrences in the high-throughput sequencing.

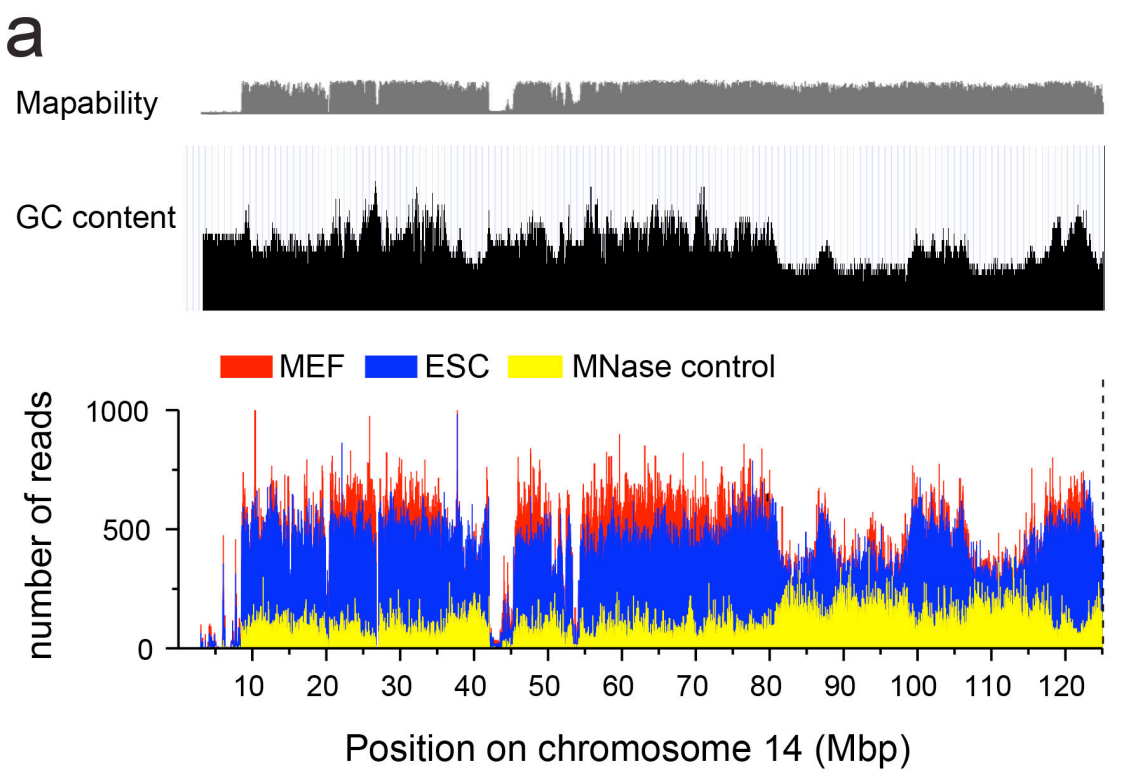
Expression profiling by RNA sequencing. Total RNA was purified and prepared for sequencing as described previously[58]. Sequencing was performed on Illumina platforms. RNA reads were aligned with TopHat[59]. Further expression analysis was performed with the Genomatix software using the most recent Eldorado gene annotation[55]. For each transcript, a normalized expression value was calculated from the read distribution that accounts for the length differences and the amount of mapped reads. The program DESeq[60] was used for the analysis of differential expression (**Supplementary Tables 6** and **7**).

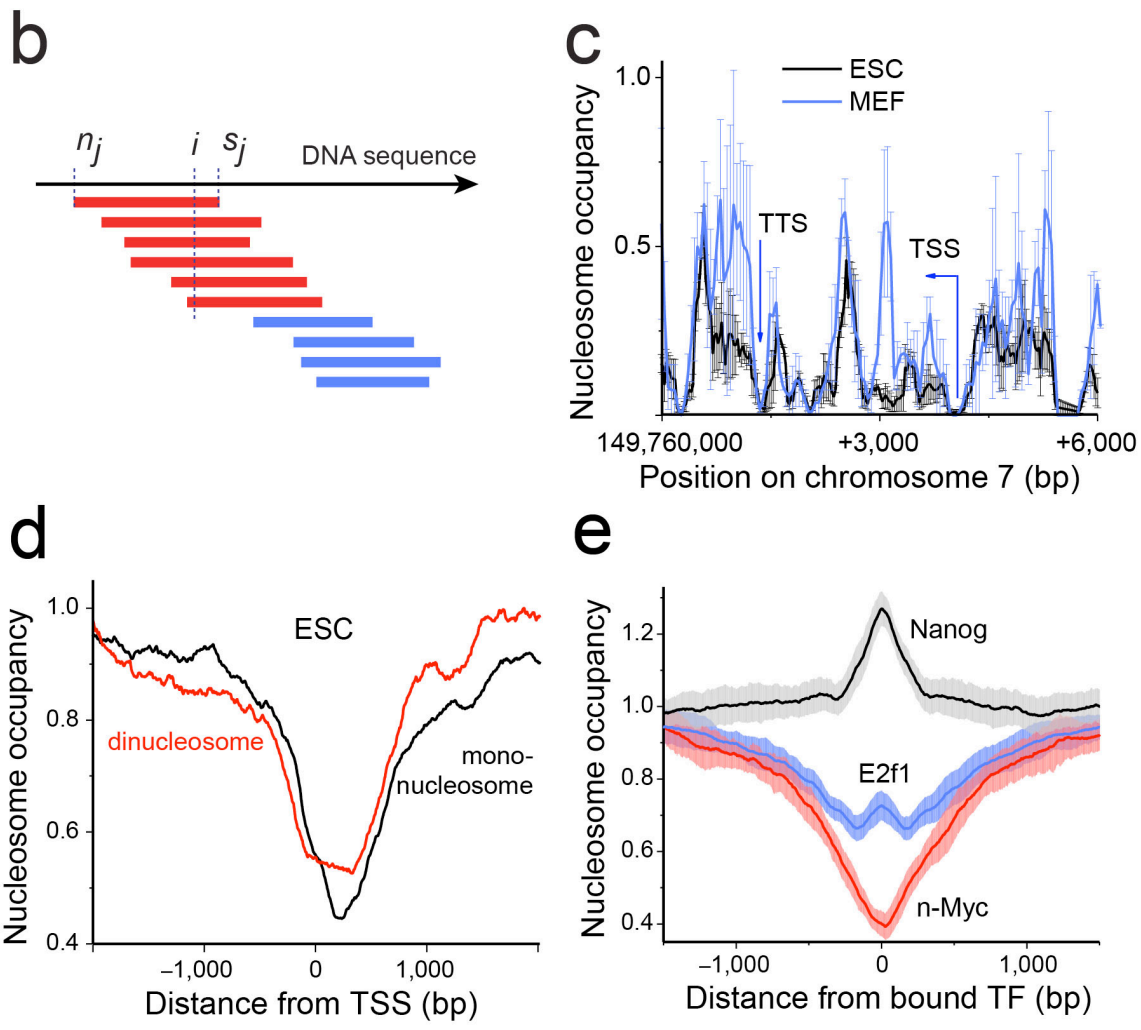
Chromatin immunoprecipitation. For each sample, 1×10^6 cells were cross-linked with 1% PFA, and cell nuclei were prepared by using a swelling buffer (25 mM HEPES, pH 7.8, 1 mM $MgCl_2$, 10 mM KCl, 0.1% NP-40, 1 mM DTT). Chromatin was sheared to mononucleosomal fragments. After IgG preclearance the sheared chromatin was incubated overnight with 4 µg of antibodies against either H3K9ac (Abcam, ab4441), H3K27ac (Abcam, ab4729) or H3K9me3 (Abcam ab8898). After washes with sonication buffer (10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5% *N*-lauroylsarcosine, 0.1% Na-deoxycholate), high-salt buffer (50 mM HEPES, pH 7.9, 500 mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS), lithium buffer (20 mM Tris-HCl, pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate) and 10 mM Tris-HCl, chromatin was eluted from the protein G magnetic beads and the cross-link was reversed overnight. After RNase A and proteinase K digestion, DNA was purified and cloned in a barcoded sequencing library for the Illumina HiSeq2000 sequencing platform. Single reads of 50-bp length were mapped with Bowtie and clustered with MACS[54], using a *P*-value cutoff of 10^{-5}.

47. Mallm, J.P., Tschape, J.A., Hick, M., Filippov, M.A. & Muller, U.C. Generation of conditional null alleles for APP and APLP2. *Genesis* **48**, 200–206 (2010).

48. Chung, H.R. *et al.* The effect of micrococcal nuclease digestion on nucleosome positioning data. *PLoS ONE* **5**, e15754 (2010).
49. Cui, P. *et al.* A novel mechanism of epigenetic regulation: nucleosome-space occupancy. *Biochem. Biophys. Res. Commun.* **391**, 884–889 (2010).
50. Allan, J., Fraser, R.M., Owen-Hughes, T. & Keszenman-Pereyra, D. Micrococcal nuclease does not substantially bias nucleosome mapping. *J. Mol. Biol.* **417**, 152–164 (2012).
51. Xi, Y., Yao, J., Chen, R., Li, W. & He, X. Nucleosome fragility reveals novel functional states of chromatin and poises genes for activation. *Genome Res.* **21**, 718–724 (2011).
52. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
53. Chereji, R.V., Tolkunov, D., Locke, G. & Morozov, A.V. Statistical mechanics of nucleosome ordering by chromatin-structure-induced two-body interactions. *Phys. Rev. E* **83**, 050903 (2011).
54. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
55. Werner, T. Next generation sequencing in functional genomics. *Brief. Bioinform.* **11**, 499–511 (2010).
56. Ward, J.H. Jr. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **48**, 236–244 (1963).
57. Huang, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
58. Caudron-Herger, M. *et al.* Coding RNAs with a non-coding function: maintenance of open chromatin structure. *Nucleus* **2**, 410–424 (2011).
59. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
60. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).

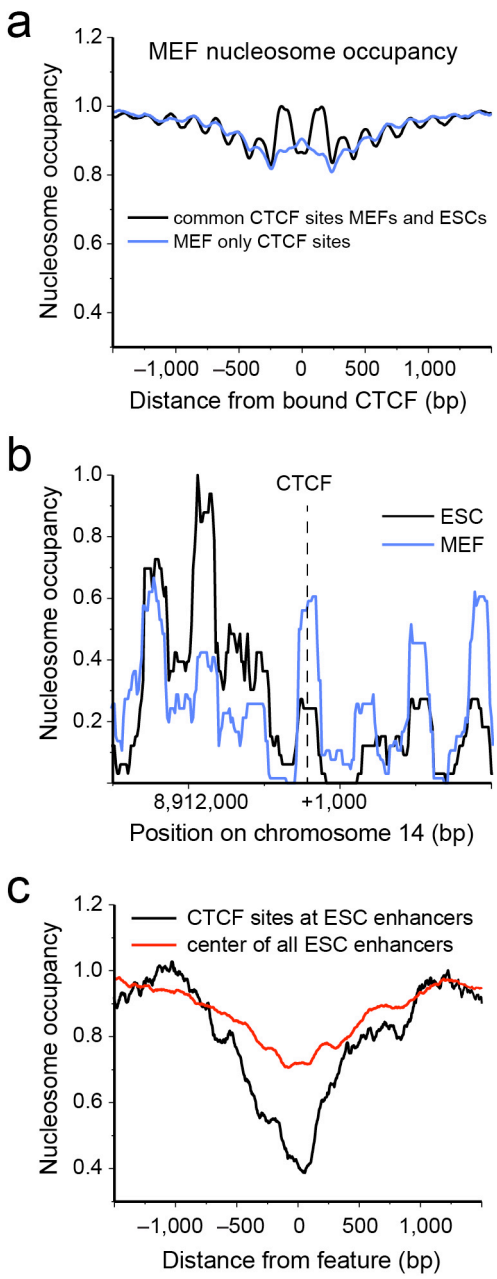
doi:10.1038/nsmb.2419


a
Mapability
GC content
MEF
ESC
MNase control
number of reads
1000
500
0
10
20
30
40
50
60
70
80
90
100
110
120
Position on chromosome 14 (Mbp)


b
n_j
i
s_j
DNA sequence
c
ESC
MEF
Nucleosome occupancy
1.0
0.5
TTS
TSS
149,760,000
+3,000
+6,000
Position on chromosome 7 (bp)
d
Nucleosome occupancy
1.0
0.8
0.6
0.4
ESC
dinucleosome
mono-
nucleosome
–1,000
0
1,000
Distance from TSS (bp)
e
Nucleosome occupancy
1.2
1.0
0.8
0.6
0.4
Nanog
E2f1
n-Myc
–1,000
0
1,000
Distance from bound TF (bp)

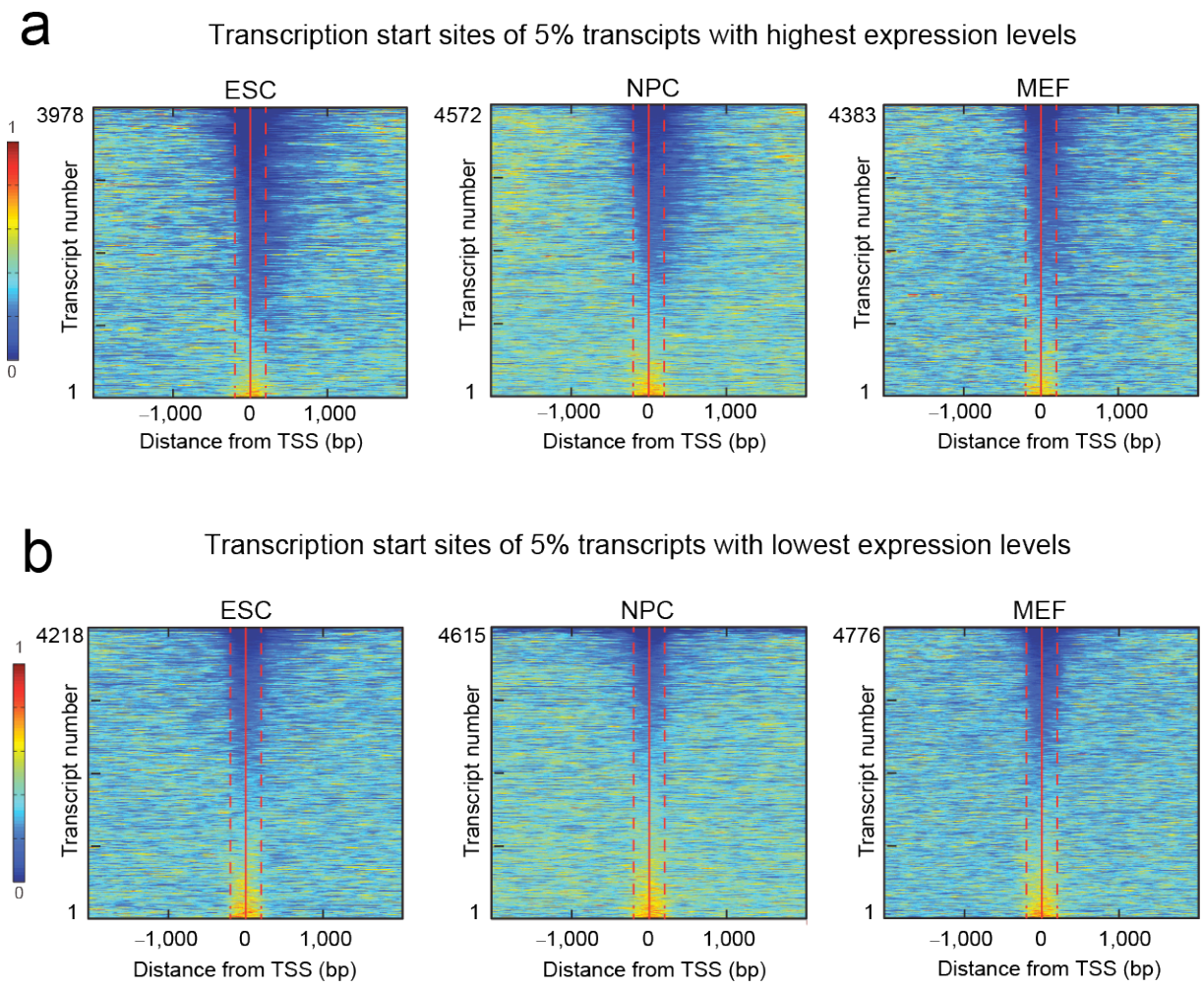
Supplementary Figure 1 Mapped sequence reads and nucleosome occupancies.

(**a**) Distribution of sequencing reads on the mouse reference genome for chromosome 14 as an example. The number of reads in a 1 Mb window was plotted for ESCs, MEFs and purified genomic mouse DNA treated with MNase. The top panel shows mapability scores for these regions as given by the UCSC Genome Browser and the average GC content for a selected region in black. The region without nucleosome occupancy signal coincided with regions for which sequence information for mapping was insufficient. The level of the nucleosome occupancy signal correlated with GC content, while the signal of the MNase control digestion of genomic DNA was anti-correlated with the GC content. (**b**) Computation of nucleosome occupancies. The paired-end sequencing data of DNA fragments obtained after MNase digestion can be directly translated into nucleosome occupancies. In our analysis the nucleosome occupancy was defined as a normalized number of individual nucleosome reads covering a given DNA position. (**c**) Accuracy of nucleosome occupancy maps for ESCs (black) and MEFs (blue) around the H19 gene. The standard errors of the average were calculated from three replicate experiments for ESCs (black) and two replicates for MEFs (blue), with each sequencing run returning ~150 million mapped reads. (**d**) Comparison of the nucleosome occupancy profiles obtained from the analysis of mononucleosome (150 – 160 bp average length) versus dinucleosome (340 bp average length) samples around the transcription start site (TSS) in ESCs for the 5% highest-expressed genes. (**e**) Accuracy of nucleosome occupancy determination in the average genome-wide profiles around TF binding sites. The nucleosome profiles of Nanog, E2f1 and n-Myc from Fig. 2 are shown together with the standard deviation at each base pair positions. The standard deviation was calculated by averaging the corresponding value for each of the mouse chromosomes.

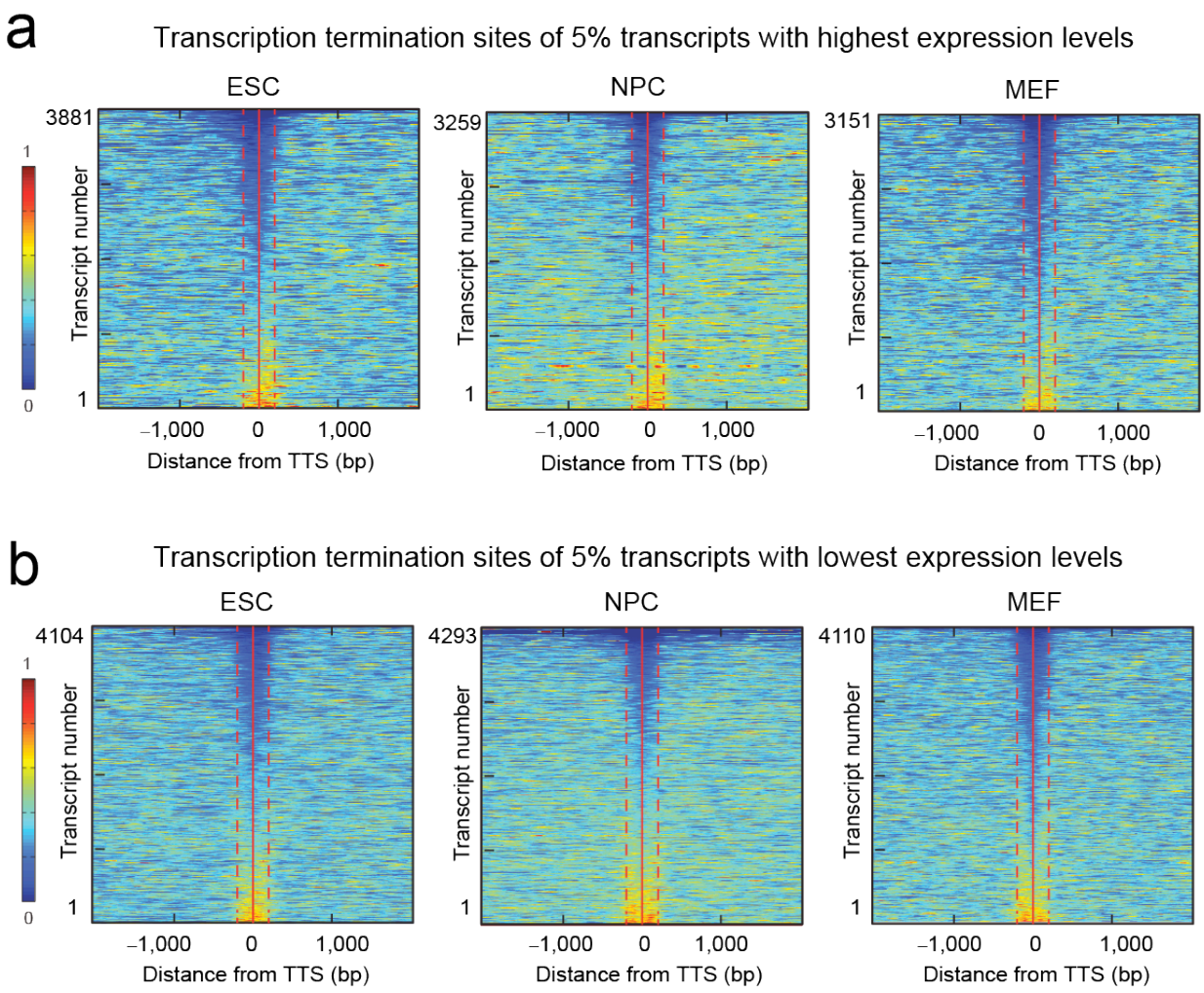


Supplementary Figure 2 Nucleosome patterns around sub-classes of CTCF sites.

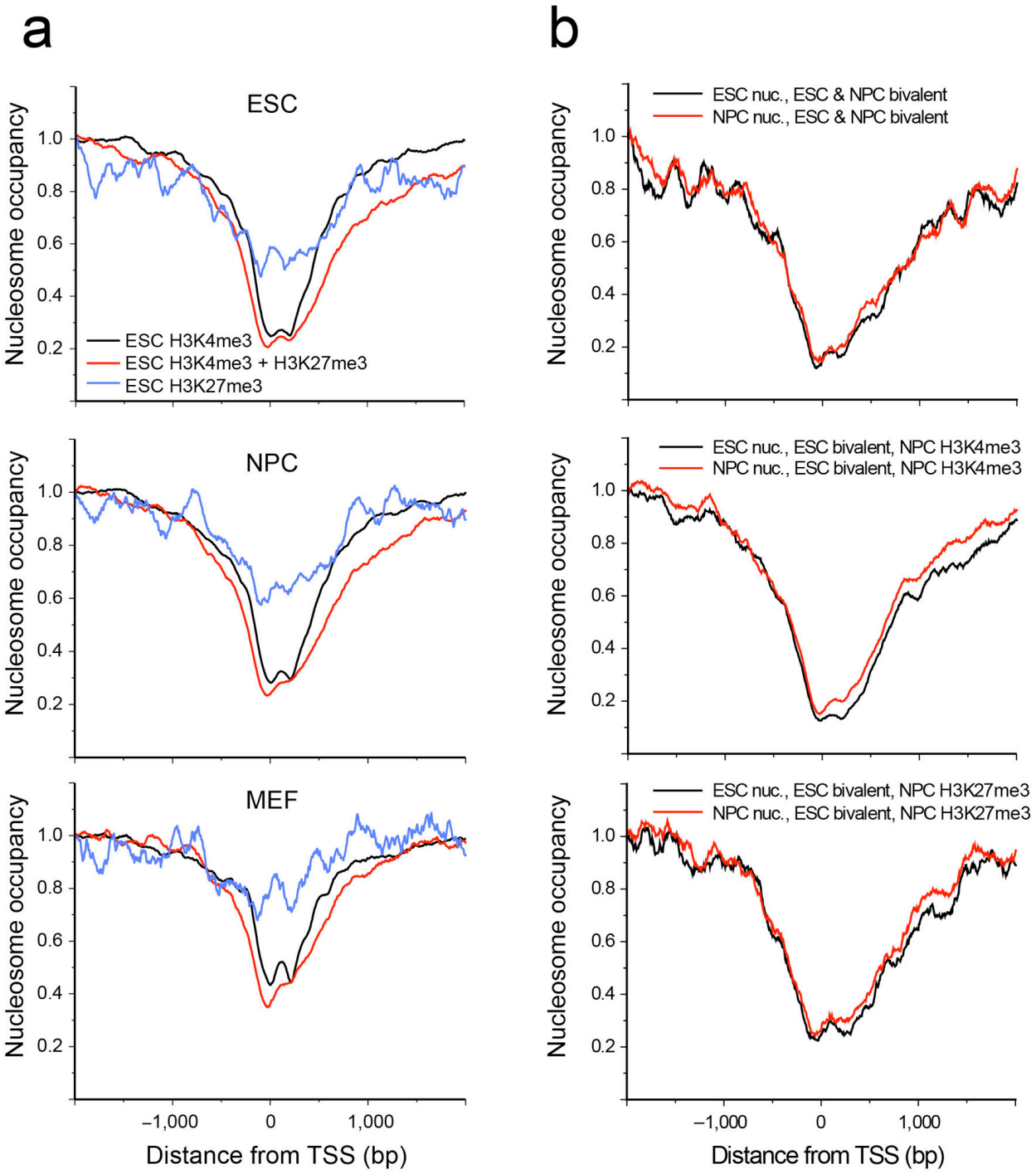
(**a**) Average patterns of MEF nucleosomes around common CTCF sites present both in ESCs and MEFs (black) are compared to average patterns of CTCF sites in MEFs that are not occupied by CTCF in ESCs (blue). (**b**) Example for a CTCF site that resides within a sequence occupied by a nucleosome in both ESCs and MEFs. (**c**) Average nucleosome occupancy patterns around common CTCF sites overlapping with putative enhancers (black) and nucleosome patterns around enhancer centers (red). The enhancer regions have been defined according to the "multiple transcription factor loci" identified previously (Chen et al., 2008, *Cell* **133**, 1106-1117).



Supplementary Figure 3 Cluster maps of nucleosome occupancy at transcription start sites. Each horizontal line corresponds to the promoter associated with a given transcript identified in the RNA-seq analysis. The colors indicate nucleosome occupancy from low (dark blue) to high (red). Nucleosome occupancies are shown for transcription start sites of the 5% transcripts with (**a**) highest and (**b**) lowest expression levels.



Supplementary Figure 4 Cluster maps of nucleosome occupancy at transcription termination sites. Each horizontal line corresponds to the termination site associated with a given transcript identified in the RNA-seq analysis. The colors indicate nucleosome occupancy from low (dark blue) to high (red). Nucleosome occupancies are shown for transcription termination sites of the 5% transcripts with (**a**) highest and (**b**) lowest expression levels.



Supplementary Figure 5 Nucleosome occupancies at promoters with H3K4me3 and H3K27me3 histone modifications in ESCs. (**a**) Promoters marked by H3K4me3 (black), H3K27me3 (blue), and bivalent promoters carrying both of these modifications (red) in each of the three studied cell types. (**b**) Promoters indicated as bivalent in ESCs, which remain bivalent in NPCs (top panel), resolve to H3K4me3 only in NPCs (middle panel), or transform into H3K27me3 only (bottom panel).

Supplementary Table 1

Comparison of gene expression levels of CTCF and linker histone variants H1.0 and H1.7 in ESCs, NPCs and MEFs.

	Normalized RNA expression levels						
	CTCF	H1.0	H1.7	H4	CTCF/H4 (ESC = 1)	H1.0/H4 (ESC =1)	H1.7/H4 (ESC =1)
ESC	0.069	0.028	0.0025	11.2	1	1	1
NPC	0.039	0.072	0.0027	1.43	4.5	20	8.6
MEF	0.015	0.057	0.0014	3.29	0.73	6.9	1.9

Normalized expression levels of a given protein were obtained by averaging the normalized expression levels of all transcripts corresponding to this protein as obtained from the DEseq analysis using the Eldorado gene annotation integrated in Genomatix.

Supplementary Table 2

Correlation of nucleosome occupancy at the transcription start site [-500,+500] versus log_2(gene expression)

Cell type	Promoter	Corr. coeff	# Transcripts
ESC	all	-0.124	78720
MEF	all	-0.069	62467
NPC	all	-0.113	75803
ESC	HCG	-0.020	31300
MEF	HCG	-0.028	28650
ESC	LCG	-0.224	3813
MEF	LCG	-0.130	2038

The transcription start site region evaluated in terms of nucleosome occupancy comprised the [-500,+500] region around the TSS. All correlation coefficients were highly significant with p-values $<10^{-10}$.

Supplementary Table 3

Correlation of log_2(MEF occupancy/ESC occupancy) at the transcription start site versus log_2(MEF expression/ESC expression)

Promoter	Corr. coeff.	# Transcripts	p-value
all	0.0487	58950	$<10^{-10}$
HCG	0.0500	27681	0.017
LCG	0.0270	2503	0.1791
bivalent ESC	0.0205	3783	$<10^{-10}$
H3K4me3 ESC	0.0005	18744	$<10^{-10}$
H3K27me3 ESC	-0.3052	104	0.085

The transcription start site region was evaluated in terms of nucleosome occupancy in the region [-500,+500] around the TSS.

Supplementary Table 4

Correlation of nucleosome occupancy changes between ESCs and MEFs at TSSs that become silenced in MEFs or ESCs

Active transcription in ESCs and silencing in MEFs			
Promoter	**% increased occupancy in MEFs**	**% decreased occupancy in MEFs**	**# transcripts total**
all	100.00	100.00	16836
HCG	18.99	13.18	2857
LCG	10.82	11.90	1885
bivalent ESC	0.30	0.05	36
H3K4me3 ESC	0.13	0.19	25
H3K27me3 ESC	0.01	0.02	2
Active transcription in MEFs and silencing in ESCs			
Promoter	**% increased occupancy in ESCs**	**% decreased occupancy in ESCs**	**# transcripts total**
all	100.00	100.00	1188
HCG	22.02	27.00	302
LCG	7.69	10.23	112
bivalent ESC	19.36	30.46	320
H3K4me3 ESC	4.77	8.63	88
H3K27me3 ESC	4.24	0.49	20

Since the global analysis shown in Table 3 did not include transcripts that were not detected either in ESCs or MEFs (corresponding to a $\log_2$(MEF expression/ESC expression) ratio of infinity), these two groups were analyzed separately. Transcripts with non-zero RNA-seq counts in ESCs and with zero reads detected in MEFs or (or vice versa) were identified. For these transcripts the change in nucleosome occupancy at the TSS in the region [-500,+500] was evaluated as given by $\log_2$(MEF nucleosome occupancy/ESC nucleosome occupancy). The number of transcripts that showed an increased or decreased occupancy for all transcripts evaluated was set to 100% to obtain the relative fraction of the different promoter subgroups that were silenced, and either displayed an increased or decreased nucleosome occupancy.

Supplementary Note

Nucleosome positions were mapped by genome-wide sequencing of nucleosomal DNA from mouse ESCs, NPCs and MEFs after digesting the linker DNA between nucleosomes with micrococcal nuclease (MNase) as described in Methods. Our determination of the nucleosome occupancy did not involve peak calling or averaging but was based on simply counting the number of times a given base pair was covered by a sequencing read (Supplementary Fig. 1b). Furthermore, no assumptions on the length of the nucleosomal DNA had to be made to derive the nucleosome occupancy, since nucleosome boundaries were determined on both sides of the nucleosome by paired-end sequencing. Thus, our analysis evaluated the raw data in a straightforward manner. From a comparison of three independent biological replicates we conclude that the error of the nucleosome occupancy value varied between 10 and 50 % for individual nucleosomes as shown for a representative region in Supplementary Fig. 1c.

The MNase digestion conditions were selected to avoid overdigestion as reflected by the presence of subnucleosomal particles (< 145 bp DNA) that form at high degrees of MNase digestion. Nucleosome occupancies at the TSS varied slightly for fragment sizes of 150 bp, 155 bp, 160 bp and 180 bp or dinucleosomes (340 bp fragment, Supplementary Fig. 1d). For the 180 bp mononucleosome and the 340 bp dinucleosome samples more reads were needed to get similar accuracies of nucleosome positions, since less fragments mapped to the border of the nucleosome. Accordingly, we included only the 150 bp, 155 bp and 160 bp samples in our final analysis.

With respect to the standard errors associated with our measurement, three types of data sets need to be distinguished in terms of the accuracy at which we were able to determine them: (i) The averaged nucleosome occupancy profiles like those shown in Fig. 2 - 6 have only a very small error as shown for the nucleosome occupancy at binding sites for transcription factors. These originate from averaging hundreds or thousands of normalized nucleosome occupancy profiles as shown for three exemplary transcription-factor binding sites in Supplementary Fig. 1e. (ii) For an individual locus the occupancy at a given base pair position had an error of 10-50% as mentioned above (Supplementary Fig. 1d). However, this type of analysis was not applied here in a quantitative manner and the conclusions of our study are based on the analysis of averaged profiles. (iii) The accuracy at which we can determine the position of individual nucleosome, i.e. the location of a given occupancy peak along the x-axis, is estimated to be ±20 base pairs as inferred from a comparison of three different independent replicate experiments (Supplementary Fig. 1c).