

High-Resolution Mapping of Chromatin Packaging in Mouse Embryonic Stem Cells and Sperm

Benjamin R. Carone,¹ Jui-Hung Hung,^{2,3} Sarah J. Hainer,^{4,5} Min-Te Chou,² Dawn M. Carone,⁶ Zhiping Weng,³ Thomas G. Fazio,^{4,5} and Oliver J. Rando^{1,*}

¹Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA

²Institute of Bioinformatics and System Biology, National Chiao Tung University, 75 Bo-Ai Street, Hsin-Chu, Taiwan 300

³Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA 01605, USA

⁴Program in Gene Function and Expression, University of Massachusetts Medical School, Worcester, MA 01605, USA

⁵Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA 01605, USA

⁶Department of Developmental and Cell Biology, University of Massachusetts Medical School, Worcester, MA 01605, USA

*Correspondence: oliver.rando@umassmed.edu

<http://dx.doi.org/10.1016/j.devcel.2014.05.024>

SUMMARY

Mammalian embryonic stem cells (ESCs) and sperm exhibit unusual chromatin packaging that plays important roles in cellular function. Here, we extend a recently developed technique, based on deep paired-end sequencing of lightly digested chromatin, to assess footprints of nucleosomes and other DNA-binding proteins genome-wide in murine ESCs and sperm. In ESCs, we recover well-characterized features of chromatin such as promoter nucleosome depletion and further identify widespread footprints of sequence-specific DNA-binding proteins such as CTCF, which we validate in knockdown studies. We document global differences in nuclease accessibility between ESCs and sperm, finding that the majority of histone retention in sperm preferentially occurs in large gene-poor genomic regions, with only a small subset of nucleosomes being retained over promoters of developmental regulators. Finally, we describe evidence that CTCF remains associated with the genome in mature sperm, where it could play a role in organizing the sperm genome.

INTRODUCTION

Eukaryotic genomes are packaged into a nucleoprotein complex known as chromatin, whose repeating subunit consists of ~147 base pairs (bp) of DNA wrapped around an octamer of histone proteins. Nucleosomes affect essentially all DNA-templated processes; as a result, there has been a great deal of interest over the past decade in characterizing nucleosome positions across the genomes of a variety of organisms (Jiang and Pugh, 2009; Radman-Livaja and Rando, 2010). In mammals, genome-wide maps have been reported for nucleosome positions in a number of cell types and tissues, including several immune cell types (Schones et al., 2008; Valouev et al., 2011); liver (Li et al., 2012); and embryonic stem cells (ESCs), neural precursor cells, and embryonic fibroblasts (Li et al., 2012; Teif et al., 2012). These genome-wide maps reveal characteristic features that are

conserved throughout eukaryotes, such as nucleosome depletion at promoters and other regulatory elements.

Several cell types exhibit unusual chromatin states that appear to be linked to biological function. ESCs, which are pluripotent, are characterized by “hyperdynamic” chromatin in which histone proteins exchange rapidly on and off of the genome (Meshorer et al., 2006). This has been proposed to contribute to a generally permissive chromatin state in which genes important for differentiation are accessible for rapid transcriptional activation. In contrast, mammalian sperm exhibit a highly unusual chromatin state that is vastly different from that of other cell types (Ooi and Henikoff, 2007); most of the histone proteins are lost during spermatogenesis, first replaced by transition proteins, and eventually replaced by small basic proteins termed protamines. However, not all histones are lost (murine sperm retain ~2% of their histones), and recent studies on histone retention in human and mouse sperm suggest that there is a bias for promoters of genes expressed early during development to be specifically packaged in histones (Arpanahi et al., 2009; Brykczynska et al., 2010; Erkek et al., 2013; Gardiner-Garden et al., 1998; Hammoud et al., 2009). These findings contrast with several lines of evidence suggesting that histone retention in sperm primarily occurs over repeat elements; small-scale cloning of DNA released by nuclease digestion of sperm revealed primarily repeat elements such as long and short interspersed nuclear element (LINE and SINE, respectively) sequences (Pittoggi et al., 1999) and pericentric repeats (Govin et al., 2007), while immunostaining studies on mature sperm reveal colocalization of histone proteins with the repeat-enriched sperm chromocenter (Govin et al., 2007; van der Heijden et al., 2006). The discrepancies between these views of the sperm chromatin landscape remain unresolved.

In general, genome-wide nucleosome mapping relies on the characterization of the products of micrococcal nuclease (MNase) digestion of chromatin; MNase preferentially cleaves the linker DNA between nucleosomes, leaving nucleosomal DNA relatively intact as ~147 bp footprints. Genome-wide characterization of MNase digestion products has proceeded rather rapidly from early studies using ~1 kb resolution microarrays to higher resolution tiling microarrays, to the modern era of deep sequencing (Radman-Livaja and Rando, 2010). Most recently, the Kent and Henikoff groups reported a significant advance in chromatin mapping (Henikoff et al., 2011; Kent

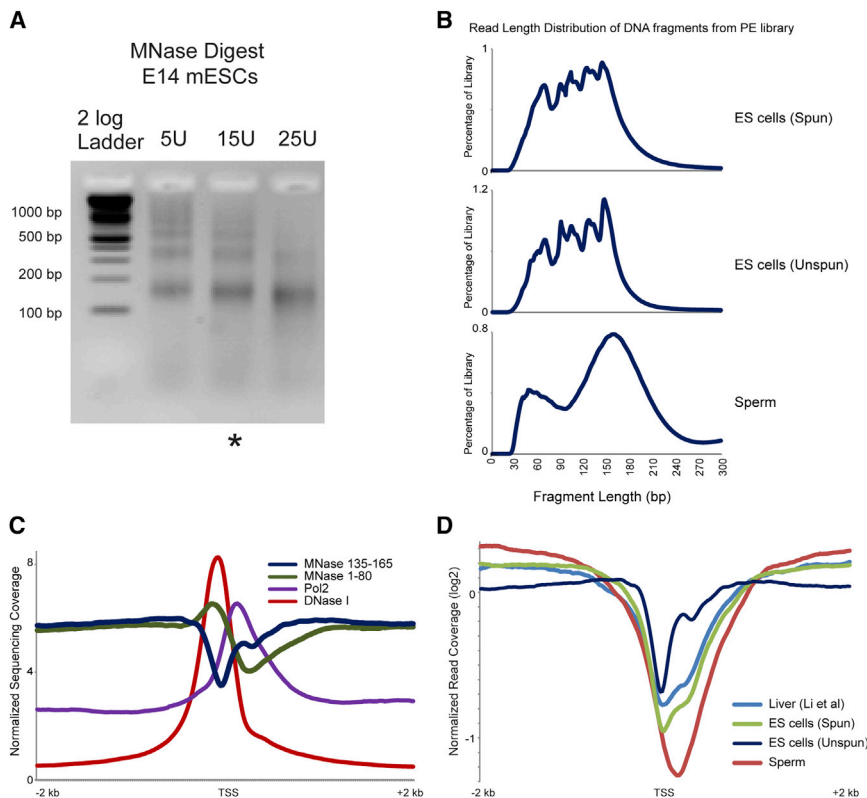


Figure 1. Average Promoter Architecture in ESCs

(A) Gel electrophoresis of MNase digestion ladders for murine ESCs subject to EGFP KD. Characteristic nucleosomal laddering is evident in underdigested ES lanes. ESC digestion patterns are shown here for unfractionated (unspun) samples; identical results are obtained for the supernatant of low-speed centrifugation ("spun," data not shown). Asterisk indicates the titration step used for paired-end deep sequencing library construction.

(B) Fragment size distribution for sequenced MNase digestion products. After deep sequencing of the indicated libraries, paired sequences were mapped back to the mouse genome, and insert size was calculated from genomic distance between reads. ES "spun" indicates that MNase digestion was fractionated by centrifugation, and the library was constructed from supernatant material, whereas the "unspun" library was constructed from the entire MNase digestion. PE, paired end.

(C) Normalized deep sequencing data plotted relative to TSSs of all mouse mm9 refGene transcripts for 1 to 80 bp MNase-protected fragments, 135 to 165 bp fragments, DNaseI hypersensitivity (GSE40869), and RNAPII ChIP-seq (indicated as Pol2), as indicated. The y axis represents the average value, for all genes, in parts per million reads.

(D) Comparison between cell types, and comparison to liver data from (Li et al., 2012). MNase

footprints (135–165 bp) were selected for the indicated libraries, and occupancy was averaged for all TSS-aligned genes, shown as \log_2 of the enrichment relative to the genome-wide average read depth. Two key features are apparent here. First, sperm cells exhibit a strong NDR on average. Second, the unspun ESC library exhibits a relatively positioned +1 nucleosome on average, whereas the supernatant "spun" library exhibits the 3'-extended promoter NDR observed in several prior studies (Teif et al., 2012; Li et al., 2012), indicating that 5' nucleosomes are relatively insoluble under typical MNase digestion conditions.

et al., 2011) by carrying out paired-end deep sequencing of an entire MNase digestion ladder (as opposed to using size-selected mononucleosomal DNA from such a ladder). These maps reveal not only nucleosome footprints of ~120–150 bp (fragments shorter than 147 bp result from MNase "nibbling" on the ends of nucleosomal DNA) but also shorter (<80 bp) footprints of other DNA-binding proteins such as transcription factors (TFs).

Here, we use this method to analyze the chromatin structure of murine ESCs and sperm. We find that nucleosome positioning in ESCs is consistent with that of many other cell types, with broadly conserved features such as promoter nucleosome depletion that scales with transcription rate. More interesting, we confirm that this protocol yields footprints of a wide variety of sequence-specific DNA-binding proteins, including pluripotency factors such as Oct4, as well as the genomic insulator CTCF. We confirm that short MNase footprints over CTCF binding sites correspond to bona fide CTCF binding events by showing that these footprints are lost on CTCF knockdown in ESCs. Furthermore, we show that sperm nuclease accessibility is broadly different than that of ESCs. In contrast to prior genome-wide studies, we demonstrate that the majority of nucleosomes in sperm are retained over gene-poor regions, and we independently confirm this result microscopically in mature sperm. Nonetheless, using a specific subset of MNase digestion conditions, we are also able to recapitulate previous reports of

histone retention at CpG-rich promoters, although we show that this signal can only account for a minority of the total levels of histone retention in sperm. Last, we unexpectedly identify widespread CTCF footprints in sperm, suggesting that this factor may play a role in organizing the sperm genome for fertilization or for embryonic genome activation.

RESULTS

We set out to thoroughly characterize the chromatin structure in murine ESCs and sperm, using a protocol recently developed by the Henikoff group (Henikoff et al., 2011). As is common in nucleosome mapping studies, cells were crosslinked with formaldehyde, lysed, and subjected to MNase digestion. Purified DNA after MNase digestion from ESCs is visualized in Figure 1A, revealing the expected nucleosome ladder. Following digestion, protected genomic DNA was isolated, and the entire nucleosomal ladder was subject to paired-end deep sequencing. In this protocol, there is no gel purification of mononucleosome-sized fragments, although several steps during sequencing library preparation select against DNA fragments longer than ~300 bp. After mapping sequencing reads back to the mouse genome, we calculated insert sizes for each fragment. Both libraries exhibited strong mononucleosome-sized protected fragments; 144 bp for ESCs and 157 bp for sperm (Figure 1B). For both libraries, subnucleosomal size peaks are also present,

with one predominant peak of ~60 bp in sperm (Montellier et al., 2013) and several smaller peaks in ESCs corresponding to 124, 104, 94, and 68 bp. The successive 20 bp reductions in peak size for ESCs likely represent the products of endonuclease cleavage at sharp bends in the nucleosomal DNA path where DNA interactions with the histone octamer are diminished (Crick and Klug, 1975; Noll and Kornberg, 1977).

We first focused on nucleosome-length (135–165 bp) MNase-protected fragments, expecting to recapitulate prior studies on ESC chromatin structure (Li et al., 2012; Teif et al., 2012). Basic characteristics of nucleosome positioning observed here are consistent with prior observations in a multitude of species, including mammalian cell types (Radman-Livaja and Rando, 2010). Promoters typically exhibit a nucleosome-depleted region (NDR) that coincides with genomic hypersensitivity to DNase I, and both of these features are centered just upstream of both the annotated transcriptional start site (TSS) and the peak of RNA polymerase II (RNAPII) chromatin immunoprecipitation (ChIP) signal (Yildirim et al., 2011) (Figure 1C; discussed later). In a prior study on ESC nucleosomes, Teif et al. reported an unusual extension of promoter nucleosome depletion downstream of the TSS (Teif et al., 2012). We do not find such a 3' NDR extension when deep sequencing unfractionated MNase-digested chromatin (Figure 1C), but this unusual NDR behavior was recapitulated in another data set in which we centrifuged the MNase digestion and isolated relatively soluble chromatin prior to generating our deep sequencing library (Figures 1B and 1D). These results suggest that nucleosomes located at the extreme 5' end of coding regions, such as the +1 nucleosome, may have reduced solubility due to their interactions with transcriptional machinery or other large protein complexes, which is consistent with prior studies on variability in nucleosomal solubility following nuclease digestion in flies (Henikoff et al., 2009). Our results using unfractionated MNase digests thus show that chromatin organization in ESCs is consistent with chromatin organization in other cell types and species, at least at the level of nucleosome positioning/occupancy.

Chromatin Organization at the TSS

In addition to nucleosome-sized DNA fragments (135–165 bp), we were also able to recover a wide range of digestion fragments, including subnucleosomal-sized DNA fragments. In yeast, such MNase footprints (<80 bp) are particularly abundant over the binding locations of sequence-specific DNA-binding proteins such as TFs (Henikoff et al., 2011; Kent et al., 2011). To extend these observations to a mammalian system, we averaged data for 1 to 80 bp fragments for all TSS-aligned genes. As shown in Figure 1C, these footprints were strongly enriched at the 5' end of the NDR, consistent with these footprints corresponding to TFs or the basal transcription machinery.

We further explored the connection between MNase footprints and promoter architecture using “V-plots” introduced by Henikoff et al. (2011). Briefly, in these plots, sequencing reads are aligned by a specific genomic position (anchor); in this case, the annotated TSSs of all genes in the mouse genome. The x axis represents the distance from the anchor along the genome, while the y axis shows the sequencing fragment size. Enrichment or depletion of footprints is shown in red or green, respectively, in a heatmap. Figure 2A shows V-plots for TSS-aligned data in

ESCs for three sets of genes grouped according to expression level. Examining nucleosome-sized DNA fragments (size range, 135–165 bp) reveals the expected behavior: nucleosomes are dramatically depleted from the promoters of highly expressed genes, with modest nucleosome depletion occurring at moderately expressed genes and minimal nucleosome depletion at silent genes (Figure 2A). At highly expressed genes, we observe a well-positioned +1 nucleosome with a slightly enlarged (~175 bp) footprint, potentially representing a nucleosome engaged either with RNAPII or with some component of the basal transcription machinery. In contrast, short DNA fragments (1–80 bp) reveal the converse behavior, with footprints being observed most strongly at promoters with the highest expression in ESCs. This is consistent with the notion that these footprints result from TFs or the basal transcription machinery, as previously suggested in flies (Teves and Henikoff, 2011). Finally, we note that intermediate-sized footprints (100–130 bp) are depleted surrounding the TSS in all three groups of genes. These footprints presumably result from MNase digestion of partially unwrapped or less stably bound nucleosomes and show little correspondence with transcription rate.

Next, we compared these TSS plots in ESCs to sperm (Figure 2B). Most dramatically, in contrast to the strong TF signal at ES promoters, sperm promoters were strongly depleted of these subnucleosomal footprints. This lack of promoter TF signal is consistent with the global, or nearly global, absence of active transcription in mature sperm (Kierszenbaum and Tres, 1975; Ward and Zalensky, 1996). Curiously, while no, or very little, transcription occurs in mouse sperm (expression values for sperm TSS plots were generated based on gene expression during the round spermatid stage), significant depletion of nucleosome-sized fragments remains evident at promoters (Figure 1D and Figure 2B). We speculate that this lack of apparent nucleosome occupancy at promoters could potentially be a remnant of transcriptional activity in round spermatids rather than the result of active repositioning or selective removal of histones during chromatin compaction in spermatogenesis (discussed later). In addition to the lack of a strong signature of the basal transcription machinery (discussed earlier), and the depletion of nucleosome footprints observed at promoters, we also noted a broadening of mononucleosomal footprint width in the sperm V-plots (Figure 2B; see also Figure 1B). The cause of this broadening is currently unclear.

The conclusions drawn from the averaged data were further confirmed on a gene-by-gene basis for ESCs. Figure 2C shows how the three size class fragments described earlier correlate with gene expression level: for each size class, all genes are ordered by ESC messenger RNA (mRNA) abundance from high (top) to low (bottom). Together, our ESC data validate our experimental approach and extend our understanding of ESC chromatin to include a broad atlas of DNA-binding proteins.

Sperm Histone Retention

Having demonstrated the utility of this protocol in ESCs, we next sought to understand the unusual chromatin state of mammalian sperm by comparing our sperm and ESC data sets in more detail. First, it is notable that, in our protocol, MNase digestion of sperm results in a nucleosome ladder very similar to that observed for somatic cells (discussed later). While this was

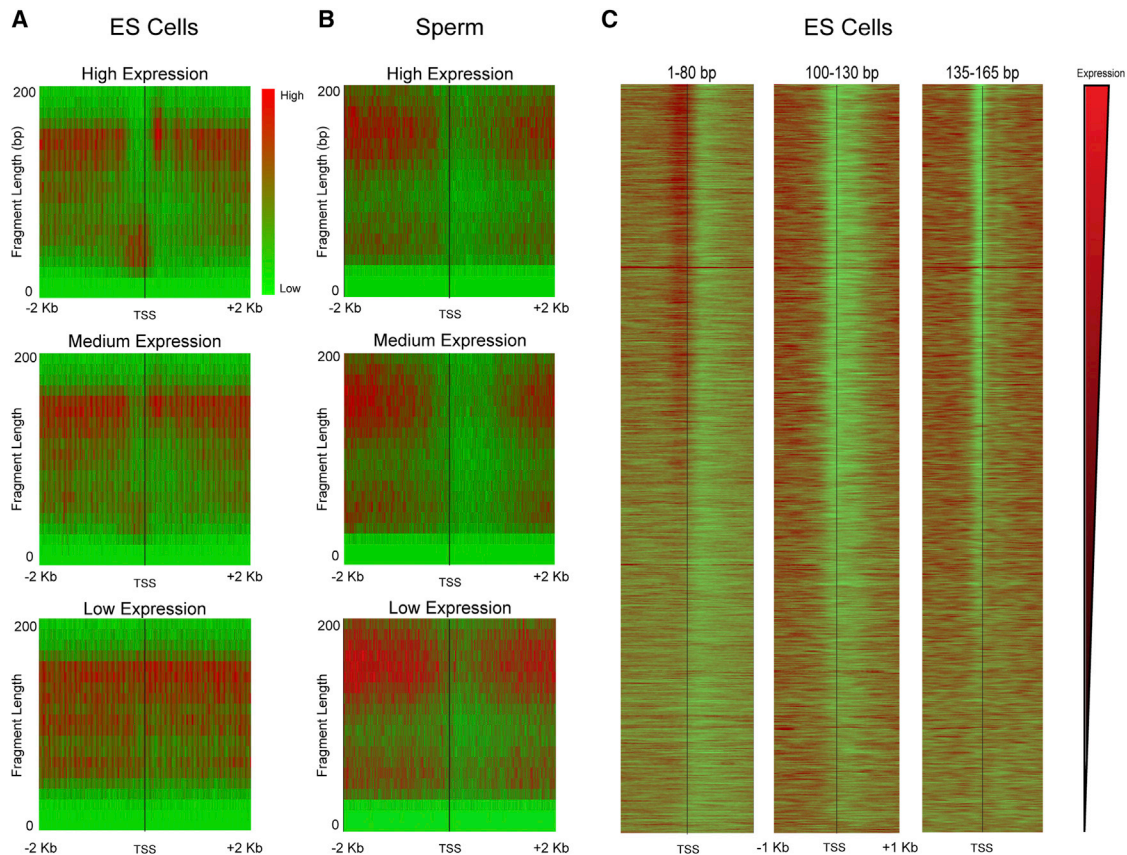


Figure 2. Relationship between Transcript Abundance and MNase Footprinting

(A and B) Enrichment of reads around TSSs is shown in a heatmap (red, enriched; green, depleted) for genes grouped according to mRNA abundance in E14 mouse ESCs: High (top 5%), Medium (middle 5%), and Low (bottom 5%). For each panel, the x axis shows distance from the TSS, and the y axis shows size of the MNase footprint. Data for ESC MNase digestion are shown in (A), and data for mature sperm are shown in (B). Note that sperm “expression levels” derive from mRNA abundance from round spermatids (Namekawa et al., 2006) as mature sperm are transcriptionally inactive.

(C) Gene-resolution heatmap of deep sequencing reads from ESC MNase digestion, aligned by TSS for three prominent size classes: <80 bp, 100–130 bp, and 135–165 bp. Each row is a single gene, and rows are sorted from high to low mRNA abundance.

expected for ESCs, the ladder observed for sperm is intriguing, given that only ~2% of histones are retained in mature mouse sperm. If sparse nucleosomes were randomly retained at 2% of genomic loci, then adjacent nucleosomes producing a dinucleosomal band would occur quite rarely; instead, the finding of a nucleosomal ladder indicates that nucleosome retention in sperm generally occurs in blocks of multiple adjacent nucleosomes.

Turning to the deep sequencing data, we noted in initial surveys of genome browser tracks that nucleosome occupancy was fairly uniform across the genome in ESCs but was substantially more variable at the megabase scale in murine sperm, with long blocks exhibiting ~2- to 3-fold variation in average nucleosome occupancy (Figures S1A and S1B available online). When comparing the enrichment of nucleosome-sized fragments (135–165 bp) between mouse ES and sperm cells using a 2 Mb sliding window, we observed that differential nucleosome occupancy was strongly correlated with gene density (Figure 3A). Specifically, nucleosomes appear to be preferentially retained in gene-poor regions in sperm relative to ESCs. To quantitate this trend genome-wide, we binned the genome into 2 Mb bins and

plotted total nucleosome-sized sequencing reads versus the number of genes within each window. As expected, ESC chromatin exhibits uniform nucleosome occupancy across the genome, whereas in sperm, the number of nucleosome-length fragments for a given window is strongly anticorrelated with the number of genes in that window (Figure 3B). Notably, the low nucleosome occupancy over promoters (Figure 1D) and gene-rich regions (Figure 3B) inferred from nucleosome-sized MNase footprints was also observed using low-coverage anti-histone H3 ChIP sequencing (ChIP-seq) (Figures S1C and S1D), validating the hypothesis that the ~150 bp MNase footprints (Figure 4A) represent bona fide histone retention in sperm.

Several previous studies examining histone retention in murine and human sperm reported strong retention of nucleosomes in the promoter regions of genes implicated in early development; these regions generally are also characterized by high CpG density and low DNA methylation (Erkek et al., 2013; Hammoud et al., 2009). This can be observed in our reanalysis of those published data sets, as TSS-aligned aggregation plots of read coverage confirms strong retention of promoter nucleosomes under the reported conditions (Figure 4B). In contrast,

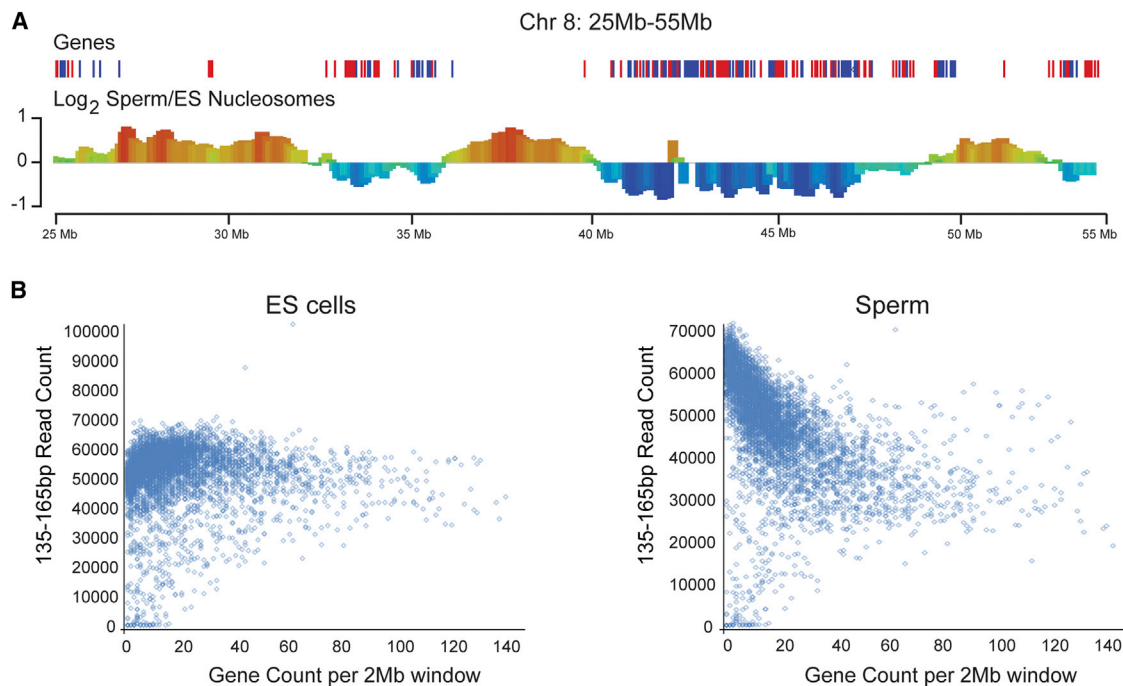


Figure 3. Preferential Histone Retention Occurs at Gene Deserts in Sperm

(A) Sperm nucleosomes are retained in gene-poor regions compared to ESC nucleosomes. Normalized mononucleosomal (135–165 bp) MNase footprints were averaged for 2 Mb bins, and the \log_2 of the relative enrichment for ESCs versus sperm (y axis) is shown for a typical stretch of chromosome 8 (genomic coordinate on x axis).

(B) Scatterplot of gene density (x axis) versus number of mononucleosome-sized MNase sequencing reads (y axis) for ESCs and sperm. The small number of points in the lower left corner depleted of sequencing reads correspond to 2 Mb bins composed of largely unmappable or unannotated sequence. Low-depth ChIP-seq using anti-H3 in sperm also reveals this anticorrelation between gene density and H3 signal, strongly supporting the claim that mononucleosome-sized fragments released by MNase digestion of sperm indeed correspond to nucleosomes.

See also Figure S1.

nucleosome-sized MNase footprints in our data set were generally depleted from promoters, including developmental promoters such as the *Hox* promoters (Figures 2B, 4C, and S1B). We speculated that these dramatic discrepancies might result from differences in the extent of MNase digestion; in yeast, several studies have reported a subset of so-called “fragile nucleosomes” that are particularly susceptible to extensive MNase digestion (Weiner et al., 2010; Xi et al., 2011). Deep sequencing of mononucleosome-sized DNA released from sperm by varying levels of MNase digestion revealed that nucleosome depletion at promoters was also observed at relatively low MNase concentrations, arguing against promoter nucleosomes being fragile (Figure 4C). In contrast, after extensive MNase digestion of sperm, we observed a ~150 bp footprint becoming enriched over promoters (Figure 4D), particularly over the “high CpG” class of promoters that are associated with developmental regulators (Figure S2A).

In addition to the extent of MNase digestion, our protocol differs from the Peters and Cairns protocols in that we digest formaldehyde-crosslinked sperm—we could not generate nucleosome-sized fragments from murine sperm in the absence of crosslinking—whereas both the Peters and Cairns groups carried out MNase digestions on uncrosslinked sperm that were first swelled in water/dithiothreitol (DTT) prior to MNase digestion. Notably, we found that, after centrifuging the sperm

MNase digests to separate primarily histone-associated DNA (supernatant) from the more abundant protamine-associated genome in the pellet, promoter retention of ~150 bp footprints was somewhat stronger in the pellet than in the supernatant (Figures S2B and S2C). Thus, we conclude that the previously reported histone retention at developmental promoters represents a specific subset of all histone retention in sperm in which promoter-bound nucleosomes are revealed only by extensive nuclease digestion—presumably, “overdigestion” eventually degrades the more abundant nucleosomes found over gene deserts, revealing the more stable promoter nucleosomes—and may furthermore be partially lost during centrifugation due to formaldehyde crosslinking to some insoluble scaffold.

Our hypothesis that histones associated with developmental promoters represent a small subset of the nucleosomes retained in sperm is also supported by elementary arithmetical considerations: human sperm retain ~10% of the nucleosome complement of a somatic cell (Hammoud et al., 2009), yet even if histones were penetrantly retained in 100% of sperm across 1 kb of sequence at all ~30,000 promoters, this would only account for 1% of the haploid genome. Nonetheless, we sought to independently assess our conclusion that most histone retention occurs in gene-poor regions in murine sperm (Figure 3B). Given the sensitivity to nuclease levels documented earlier for digestion-based assays, we used fluorescence

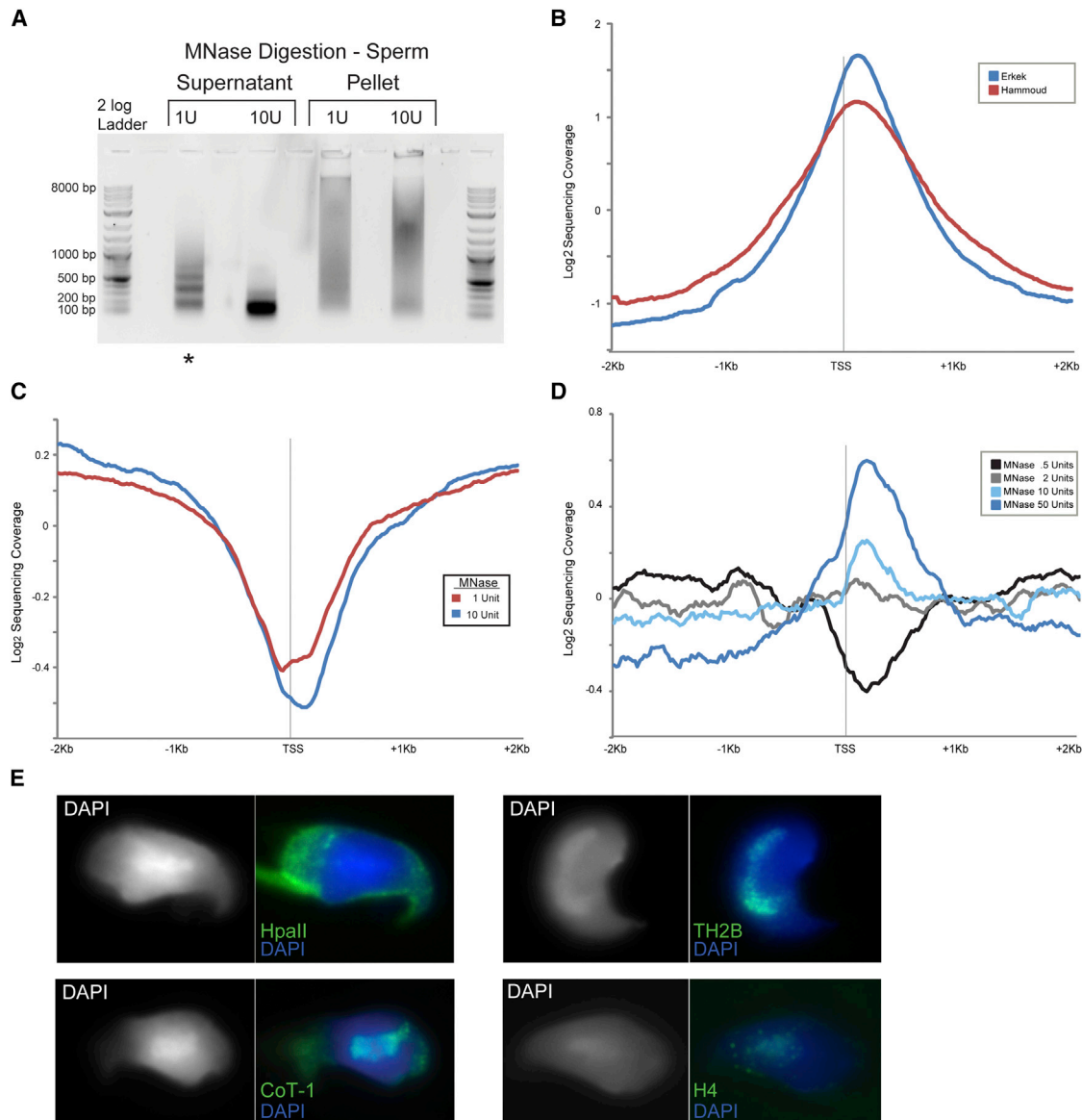


Figure 4. A Small Subset of Nucleosomes Is Retained at Promoters in Sperm

(A) Gel electrophoresis of soluble versus insoluble MNase digestion products from mature *M. musculus* sperm. Lane with asterisk designates level of digestion used for most analyses. Pellet fraction contains the majority of protamine protein, while supernatant carries the majority of histone proteins.

(B) Aggregation plot of data from Hammoud et al. (2009) (*H. sapiens*) and Erkek et al. (2013) data sets (*M. musculus*) aligned by TSS.

(C) Aggregation plot of MNase sequencing reads for 1 U and 10 U MNase digestion levels [digestions shown in (A)]. These libraries were generated from the supernatant of the MNase digestion material.

(D) Aggregation plots of MNase sequencing data for unspun, not size-selected, digestion series of formaldehyde-crosslinked *M. musculus* sperm. A peak of promoter nucleosome occupancy is revealed only after extensive MNase digestion [note that MNase levels are not directly comparable to those in (C) as fewer sperm were used in this digestion] and is more prominent when MNase digestions are not subject to centrifugation.

(E) Representative images for FISH with HpaII probe (poorly methylated CpG islands; see Experimental Procedures) and Cot-1 probe (gene-poor heterochromatic regions), as well as immunofluorescence using antibodies to TH2B (testis-specific H2B) and bulk H4, demonstrating that the majority of anti-histone staining overlaps the DAPI-rich chromocenter that also stains with Cot-1 probe.

See also Figure S2.

in situ hybridization (FISH) and immunofluorescence as a completely independent assay for histone localization in mature mouse spermatozoa. Mammalian sperm cells contain a highly organized nucleus with a DAPI-dense heterochromatic center surrounded by a euchromatic periphery that exhibits more

modest DAPI staining (Li et al., 2008; van der Heijden et al., 2006). To determine the location within the sperm nucleus of CpG-rich developmental promoters, we generated fluorescent probes from 50 to 500 bp size-selected DNA fragments of HpaII-digested genomic DNA. Cot-1 DNA probes were used to

confirm the physical location of repetitive heterochromatic sequences. Signal from the HpaII probe is prominent in the nuclear periphery, while Cot1 signal overlaps almost perfectly with the chromocenter, as expected (Figure 4E). We then used immunofluorescence to characterize the localization of histones H4 and TH2B, finding that the majority of histone staining occurs in the central DAPI-dense chromocenter, which is consistent with prior studies (Figure 4E) (Govin et al., 2007; Meyer-Ficca et al., 2013; van der Heijden et al., 2006). Taken together, our data confirm that a small subset of histone retention occurs at CpG-rich promoters but demonstrate that the vast majority of histone retention in mammalian sperm occurs in gene deserts.

Short MNase-Protected Fragments Coincide with DNA-Binding Protein Footprints

We finally turn to more detailed analysis of short MNase footprints in sperm and ESCs. As shown elsewhere (Kent et al., 2011), short MNase-protected footprints can reveal binding sites for DNA-binding proteins such as TFs. In order to determine whether our data revealed the locations of specific DNA-binding proteins, we sought sequence motifs enriched in short MNase-protected fragments relative to the background of longer (nucleosome or dinucleosome) footprints (see [Experimental Procedures](#)). In ESCs, significantly enriched motifs included sequence matches for a number of TFs that play key roles in the pluripotency transcriptional network, such as Klf4, as well as a strong match to the binding site for key insulator and locus control regulator CTCF. Plotting short MNase footprints around in vivo binding sites for a variety of TFs reveals a range of footprinting behaviors for TFs (Figure S3). Despite the lack of active transcription in sperm, we were also able to identify a number of significantly enriched motifs among the short MNase footprints in sperm. Sperm footprints did not carry strong matches to the Klf4 binding site or most other pluripotency TFs but instead were enriched for a subset of sequence motifs from the ESC motif list such as Stat3, RelA, and, most notably, CTCF (Figures 5A and S3). In contrast, a number of TFs, including key pluripotency factors Oct4, Nanog, Sox2, and Klf4, were associated with strong 1 to 80 bp footprints specifically in ESCs but not in sperm. Notably, we did not identify any motifs specifically protected in sperm but not in ESCs.

To further explore MNase footprinting patterns around TF binding sites, we focused on Klf4 and CTCF, both of which have previously been mapped in murine ESCs (Chen et al., 2008). Using matches to the relevant sequence motif located within previously identified ChIP-seq peaks as anchors ($n = 15,657$ for CTCF, $n = 1,813$ for Klf4), we generated V-plots for ESC MNase footprints (Figure 5B). As previously shown for yeast general regulatory factors such as Reb1 (Henikoff et al., 2011), three characteristics are apparent in these plots. First, strong enrichment in the 30 to 60 bp size range is observed centered on the TF motif, indicating presumed protection of DNA from MNase by the bound TF. Second, a V-shaped depletion of fragments results from MNase cleavage immediately adjacent to the TF. Finally, enriched 140 to 160 bp fragments located outside this V are consistent with well-positioned nucleosomes occurring flanking the TF binding sites, as previously reported for CTCF (Fu et al., 2008; Teif et al., 2012). Similar, albeit weaker, results were observed over Klf4 anchors.

In contrast to the extensive literature on ESC TF binding, little is known about the localization, or even presence, of sequence-specific DNA-binding factors in mature sperm. However, as noted earlier, we identified a number of sequence motifs enriched among short MNase footprints in mature sperm, most notably, the CTCF motif (Figure 5A). We repeated the V-plot analyses for CTCF and Klf4 binding sites for sperm MNase data (Figure 5C). Consistent with our motif searches, we found little evidence for Klf4 binding in mature sperm, but we confirm enrichment for <60 bp fragments at CTCF binding sites (Figures 5C and 5E). The modest enrichment observed here relative to the ESC peak is explained by the fact that only a subset (~40%) of the CTCF motifs that are bound in ESCs are associated with MNase footprints in mature sperm (Figure 5D).

Subnucleosomal MNase Footprints Represent Bona Fide CTCF Binding Events

The presence of sequence motifs underlying short MNase footprints provides us with strong hypotheses regarding TF binding events, which make several additional testable predictions. First, the extent of footprinting over CTCF motifs should correlate with in vivo occupancy levels as assayed by ChIP-seq. Second, these footprints should be eliminated by CTCF knockdown (KD). To test these predictions, we carried out ChIP quantitative PCR for six loci predicted to be CTCF binding sites in ESCs, as well as three CTCF motifs only exhibiting MNase footprints in sperm. All nine loci exhibited the expected CTCF binding in ESCs, and this binding was lost on CTCF KD (Figure S4A). More globally, we made use of published CTCF ChIP-seq localization data generated using ESCs (Chen et al., 2008). Enrichment of short MNase fragments over CTCF binding motifs was indeed correlated with in vivo CTCF occupancy, consistent with the idea that these short footprints correspond to CTCF protected fragments (Figure 6A). It is interesting that nucleosome phasing surrounding CTCF binding sites also correlated with CTCF binding levels (Figure 6B), supporting the hypothesis that CTCF acts as a positioning barrier to establish nucleosome arrays (Radman-Livaja and Rando, 2010).

Finally, we directly assessed whether short MNase footprints over CTCF binding sites resulted from bona fide CTCF binding (as opposed to MNase sequence biases or binding by a factor other than CTCF, such as its paralog CTCFL/BORIS) by repeating our MNase mapping experiment in ESCs following CTCF KD. Overall, we achieved a ~90% KD of CTCF as illustrated by both western blotting and mRNA abundance (Figures 6C and 6D). Genome-wide, results for this experiment were strongly correlated with results from control or enhanced green fluorescent protein (EGFP) knockdown ESCs, with similar fragment size distributions, nucleosome depletion at promoters, and short transcription-correlated promoter footprints (data not shown). However, as predicted, short fragments over CTCF binding sites were greatly diminished in the CTCF KD cells (Figure 6F; Figures S4B and S4C). The remaining short fragments could result either from residual CTCF following incomplete (~90%) knockdown or from the CTCFL/BORIS protein that binds a similar sequence motif to CTCF (Jelinic et al., 2006). Nucleosome phasing was also lost surrounding CTCF binding motifs in the knockdown cells (Figure 6F), providing strong evidence that CTCF binding (rather than some aspect of DNA

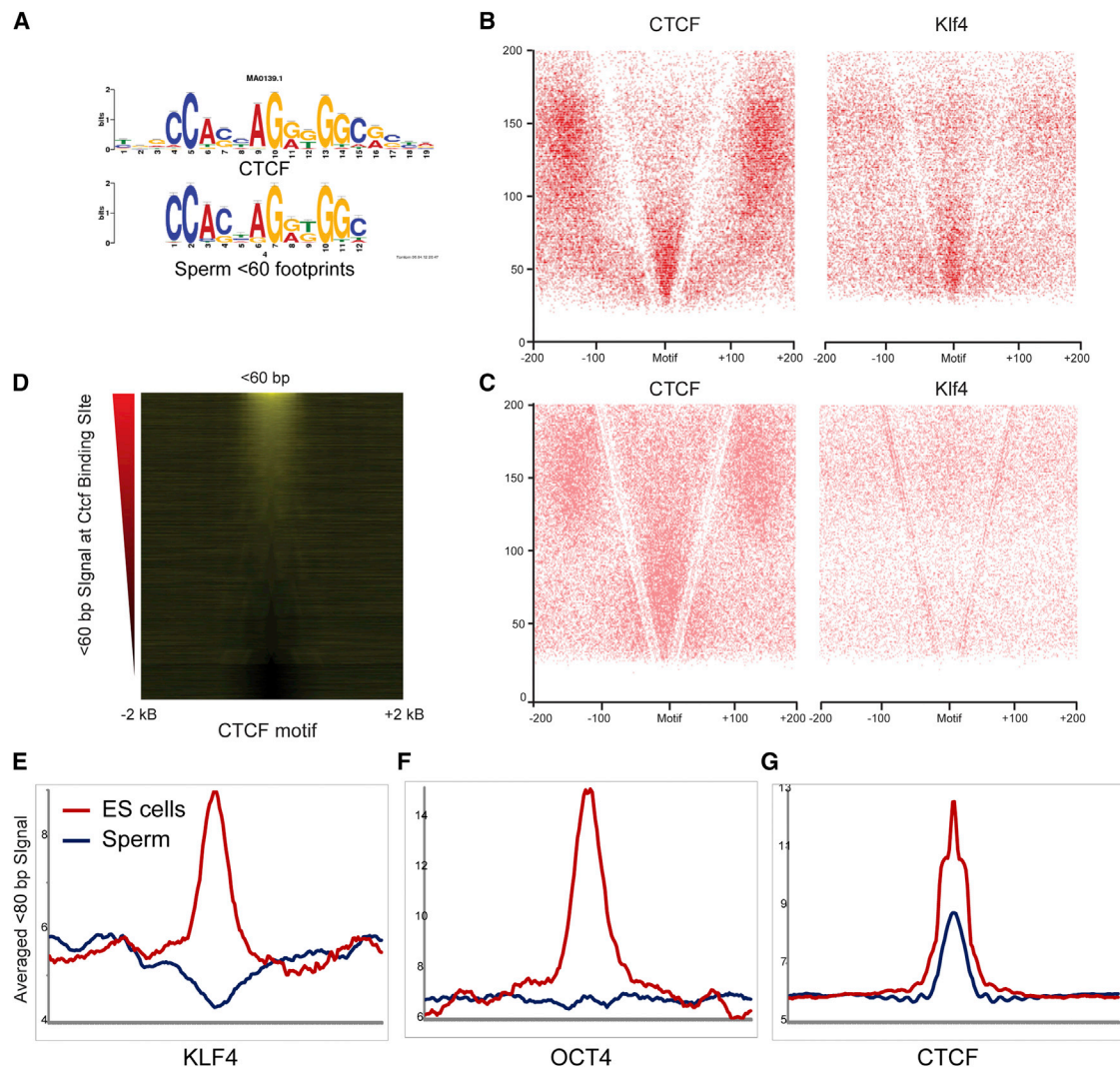


Figure 5. Enrichment of Short MNase Fragments over Binding Sites for Sequence-Specific TFs

(A) Short MNase footprints in sperm occur over CTCF binding sites. MNase digestion products (<60 bp) from sperm were searched for overrepresented sequence motifs, and the top motif hit is shown alongside the published sequence motif for CTCF.

(B and C) V-plots (as in Figure 2A, but showing all digestion products as points rather than enrichment/depletion) for ESC (B) or sperm (C) MNase digestion products, anchored by CTCF and Klf4 binding sites.

(D) Short footprints in sperm are associated with a subset of the CTCF binding sites identified in ESCs. Heatmap shows enrichment of <60 bp DNA fragments from a sperm data set, centered on CTCF motifs that are empirically CTCF-bound in ESCs (Chen et al., 2008). Sites are sorted from high to low signal intensity over 100 bp surrounding the binding site.

(E) Examples of short (<80 bp) footprints averaged over various TF motifs.

See also Figure S3.

"programming" of chromatin structure) is required for nucleosome positioning at these locations.

Finally, we asked whether MNase footprints in sperm were likely to result from CTCF itself or from CTCFL/BORIS, which binds to the same sequence and plays important roles in male reproduction (Jelinic et al., 2006; Loukinov et al., 2002). First, western blotting confirms the presence of CTCF protein in mature sperm (Figure S4D), which is consistent with prior reports (Tang and Chen, 2006). Despite the presence of this protein in sperm, to date, we have been unable to successfully recover ChIP material; this is likely due to the unique crosslinking and fragmentation

challenges presented by the highly compact sperm nucleus. As an indirect approach to the question of the CTCF footprinting protein, we focused on CTCF-specific and BORIS-specific binding events recently described in ESCs (Sleutels et al., 2012) and asked whether these sites were occupied by short MNase DNA fragments from our sperm data set. We find that <60 bp MNase footprints over CTCF motifs in sperm preferentially occur over sites bound exclusively (in ESCs) to CTCF, while they are completely absent over known CTCFL-only binding sites (Figure S5). These results support our hypothesis that CTCF itself is associated with the genome in mature murine sperm.

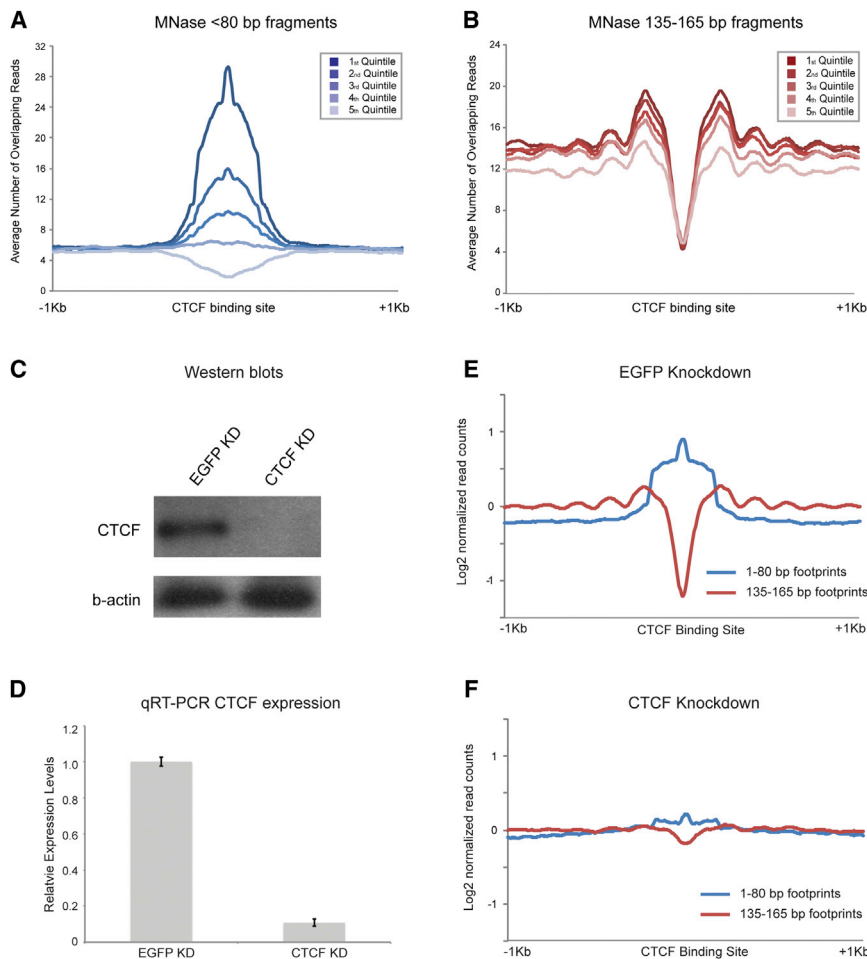


Figure 6. Nuclease-Resistant Footprints over CTCF Motifs Represent Bona Fide CTCF Binding in ESCs

(A) Abundance of short MNase footprints over CTCF binding sites in ESCs correlates with CTCF ChIP-seq enrichment. CTCF motifs are split into quintiles according to ChIP-seq signal (Chen et al., 2008).

(B) As in (A), but for mononucleosome-length footprints. Flanking nucleosomes are more strongly positioned when CTCF ChIP-seq signal is highest.

(C) Western blots of CTCF and EGFP esiRNA KD in mouse ESCs, probed with anti-CTCF and β -actin antibodies.

(D) Quantitative RT-PCR of CTCF mRNA abundance.

(E) Aggregation plot for <80 bp and 135–165 bp digestion products in EGFP KD ESCs, aligned using CTCF motifs.

(F) KD of CTCF in ESCs results in loss of <80 bp footprint enrichment over CTCF motifs, with an associated increase in nucleosome occupancy over the CTCF motif and loss of surrounding nucleosome positioning.

See also Figures S4 and S5.

footprinting at experimentally validated binding sites for all TFs (Figure S3). This was previously observed in yeast studies, in which an unusual subset of TFs (largely corresponding to the abundant “general regulatory factors”) showed far stronger footprinting than the majority of TFs (Henikoff et al., 2011). The absence of strong MNase

footprinting at certain TF binding sites (such as for Esrrb in ESCs) could reflect weak crosslinking to DNA, low average occupancy, or inability of the DNA-bound TF to interfere with MNase digestion. It is interesting in this regard that the binding sites associated with strong footprints, such as CTCF and Klf4 binding sites, tended to be associated with relatively well-positioned flanking nucleosomes (Figure 5B). One potential explanation for this correlation would be that TF footprints are strongest for factors associated with particular chromatin remodeling complexes.

We tested the hypothesis that these short MNase footprints report on bona fide TF binding in vivo, finding that CTCF footprints exhibited strong correlation with CTCF ChIP-seq data and further showing that MNase footprints over CTCF binding sites are greatly reduced on CTCF KD. Finally, we extended the prior observation that CTCF binding is associated with strong nucleosome phasing (Fu et al., 2008) by showing that the level of CTCF binding corresponds to the strength of nucleosome positioning, supporting the hypothesis that TFs can function as positioning barriers. Moreover, loss of CTCF binding on KD leads to a loss of nucleosome phasing around this site. Taken together, these results lend further support to the hypothesis that TFs are capable of directing nucleosome positioning.

DISCUSSION

By characterizing a wide size spectrum of MNase digestion products, we have profiled the chromatin architecture of two unusual mouse cell types, ESCs and sperm, illuminating nucleosome occupancy and TF binding in each cell type and providing strong evidence for the presence of bound TFs such as CTCF in mature sperm.

ESC Chromatin

In ESCs, our analysis of mononucleosome-sized footprints confirms well-known aspects of chromatin biology, such as a strong relationship between promoter nucleosome depletion and transcription level. We therefore focused on analysis of features only appreciable in non-size-selected libraries. Extending our analysis to short MNase footprints reveals a strong correlation between these footprints at the 5' ends of promoters and transcription rate, thus supporting the hypothesis (Teves and Henikoff, 2011) that these MNase footprints reflect the presence of the basal transcription machinery (among other DNA-binding proteins).

Short MNase footprints were enriched for sequence motifs associated with a subset of key pluripotency regulators, such as Klf4. It is interesting that we did not observe equally strong

Analysis of Sperm Histone Retention

The unusual packaging state of mammalian sperm, which retain ~2%–15% of the histone load of somatic cells (Ooi and Henikoff, 2007), has been proposed to play roles in programming early embryonic gene expression (Brykczynska et al., 2010; Gardiner-Garden et al., 1998; Hammoud et al., 2009) and in epigenetic inheritance of cytosine methylation states (Nakamura et al., 2012). The specialized subset of genomic loci associated with retained histones in sperm is, therefore, of great interest.

In principle, retention of nucleosomes over 2% of the genome in sperm could occur stochastically (e.g., each sperm cell retains nucleosomes over a different 2% of the genome), deterministically (all sperm cells retain nucleosomes over specific loci), or somewhere in between. Prior reports on sperm histone retention in humans and mice, using nuclease sensitivity as a key assay (Arpanahi et al., 2009; Brykczynska et al., 2010; Hammoud et al., 2009), came to somewhat different conclusions. Two studies isolated solubilized mononucleosome-sized fragments after MNase digestion of human sperm, finding enrichment over the promoters of key developmental regulators, with a very strong correlation with GC content (Brykczynska et al., 2010; Hammoud et al., 2009; Vavouri and Lehner, 2011). Somewhat different results were obtained via tiling microarray analysis of DNA released after light restriction enzyme or MNase digestion of human and mouse sperm, with gene and promoter-rich regions flanked by CTCF sites very broadly overrepresented in much longer (megabase-scale) domains (Arpanahi et al., 2009; Saïda et al., 2011).

Here, we confirm that nucleosome-sized fragments can be found over CpG-rich promoters but find that these nucleosomes are unusual in that they are only released by extensive MNase digestion and appear to be crosslinked by formaldehyde into a somewhat insoluble complex. Moreover, we show that these nucleosomes represent a small fraction of all nucleosomes retained in mature sperm; preferential histone retention over gene deserts is observed robustly across a range of MNase digestion levels, and fluorescence microscopy independently confirms that the majority of histone protein is localized in a distinct chromocenter occupied by repeat elements rather than the CpG-rich fraction of the genome associated with developmental promoters (Figure 4E). Whether either of these classes of histone-bound loci serves any function in early embryonic processes (see Ihara et al., 2014)—and how these functions might differ between the broad blocks of nucleosomes occurring over repeat elements (e.g., gene deserts) and the more focal nucleosomes found over CpG islands—is currently unclear. Notably, recent studies have shown that retention of H3K9me2-marked histones plays a role in protecting a fraction of the paternal genome from Tet3-dependent cytosine demethylation (Nakamura et al., 2012). Preliminary low-coverage anti-H3K9me2 ChIP-seq shows preferential localization of this mark to gene-poor regions of the genome (data not shown). Given that we found histone retention primarily occurring over gene deserts, our results suggest a model in which the sperm packaging state preferentially protects epigenetic silencing marks from removal at repeat elements. Advances in editing the epigenome should allow targeting histone eviction or covalent modifications at specific loci in gametes to definitively determine the role for sperm histones in zygotic biology.

Finally, we provide evidence that a subset of CTCF binding sites is associated with strong MNase footprints in mature sperm. This suggests the hypothesis that mature sperm carry chromatin-associated CTCF at specific genomic locations. CTCF protein has previously been reported to be present in mature mouse sperm (Tang and Chen, 2006), although it was not shown to associate with the sperm genome. CTCF is intricately involved in epigenetic control of a number of well-studied imprinted genes (Phillips and Corces, 2009), raising the intriguing possibility that CTCF associated with the paternal genome plays a regulatory role in gene expression during early development. Future studies using conditional alleles of CTCF will be needed to assess if and how chromatin-bound CTCF protein in sperm influences gene expression on zygotic genome activation and whether inheritance of CTCF-bound sperm nuclei is essential for developmental gene regulation.

EXPERIMENTAL PROCEDURES

Tissue Culture and Sperm Isolation

ESC lines used in this study were mouse E14 lines cultured using standard conditions. Briefly, E14 ESCs were cultured on gelatin-coated dishes without feeder cells in standard media containing serum and leukemia inhibitory factor at 37°C with 5% CO₂. ESCs were passed by washing with PBS and dissociating with trypsin. Sperm cells were isolated from the caudal epididymis and vas deferens of 10-week-old C57/6J mice by lacerating tissue in 1 ml of 37°C M2 medium (Sigma). Sperm cells were allowed to swim up for 1 hr, washed 1× in water, 1× PBS, crosslinked at 37°C with 1% formaldehyde for 15 min, quenched with 250 mM glycine for 5 min, and frozen in liquid nitrogen. The majority of figures were generated using data from control EGFP KD ES cultures, with the exception of Figures 4E and S3, which used untreated ESC data that had higher levels of short MNase footprints. Control and EGFP KD data yielded nearly identical results; EGFP KDs were used simply based on the higher sequencing depth for these libraries. Animals were maintained on site in accordance with approved Institutional Animal Care and Use Committee protocols.

Sperm Permeabilization

Frozen sperm cells were briefly thawed at 37°C, pelleted 4,000 × g for 4 min at room temperature (RT), and washed 1× in 1 ml PBS; then, the pellet was resuspended in 1 ml sperm permeabilization buffer (SPB) (0.025% trypsin/EDTA; 10 mM DTT; 5 mM Tris, pH 7.5; 5 mM NaCl; 1.5 mM MgCl₂; 0.25% NP40; 1 mM phenylmethylsulfonyl fluoride [PMSF]). Sperm cells were incubated in SPB for 5 min at 37°C. After 5 min, 10 μl of fetal bovine serum was added to inactivate trypsin, and the sample was incubated for an additional 5 min at 37°C. Sperm cells were pelleted 4,000 × g for 4 min at RT, washed with PBS 2×, resuspended in 500 μl PBS + 2 mM CaCl₂, and incubated 5 min at 37°C.

RNAi-Mediated KD

RNAi using endoribonuclease III small interfering RNAs (esiRNAs) was performed as previously described (Fazio et al., 2008), transfecting E14 mouse ESCs using Lipofectamine 2000. Briefly, templates for esiRNA production were produced using a two-step PCR procedure followed by in vitro transcription and dicing of double-stranded RNA using RNase III. KDs were performed for 48 hr to achieve effective KD.

MNase Digestion

Permeabilized ES and sperm nuclei were treated with 10 U per 10⁶ cells and 1 U per 10⁸ cells of MNase (Worthington), respectively, for 5 min at 37°C. Reaction was stopped with the addition of 10 mM EGTA on ice. For ESCs, nuclei were incubated for 4 hr at 4°C with rotation, pelleted at 5,000 × g for 5 min, and separated into supernatant and pellet (spun) or were treated the same without centrifugation and separation (unspun). After MNase digestion of sperm nuclei, digestion was centrifuged for 5,000 × g for 5 min: the

supernatant from this spin contained the vast majority of histone protein, whereas the vast majority of protamine protein was located in the pellet (data not shown). All sperm data except in Figures 4D, S2A, and S2C were generated using the supernatant fraction; the sperm data in Figures 4D and S2A were generated using libraries constructed without any centrifugation step. For Figures S2B and S2C, libraries were prepared from sperm digested with a range of MNase concentrations—1, 5, 20, and 100 U—for 5 min at 37°C, where digestion was stopped with EGTA on ice, and incubated for 4 hr at 4°C. Supernatant and pellet fractions were separated by centrifugation at 5,000 × g for 5 min, and the ~150 bp mononucleosome-sized DNA was used to generate deep sequencing libraries. In all cases, DNA was isolated by phenol:chloroform:isoamyl alcohol and EtOH precipitation.

Sperm ChIP-Seq

Mouse sperm cells were isolated as described earlier. Chromatin isolated from six mice was used for the immunoprecipitation experiments. Crosslinked sperm cells were washed with 1× with PBS, resuspended in 600 µl of 20 mM Tris with 20% glycerol and 10 mM PMSF and subjected to bead-beating (0.5 mm glass beads). SDS was added to a final concentration of 1%, and chromatin was sheared to ~300 bp with Bioruptor. Samples were kept on ice at all stages. ChIP was performed overnight at 4°C with Pan-H3 antibody (abcam, ab1791).

Deep Sequencing

To characterize the entire range of MNase digestion products, we generated paired-end libraries of DNA fragments from both ES and sperm digestions and performed sequencing using Illumina HiSeq technology. Libraries were prepared as described by Henikoff et al. (2011). Reads were mapped to the mouse genome (mm9) using Bowtie2, and uniquely mapped reads were used for further analysis. For library insert size distributions, see Figure 1B. For most figures, reads were normalized to parts per million reads. In Figures 4, 6, and S2, data are normalized to the average genome-wide coverage and represented as log₂ of the fold difference from the genomic average.

The majority of analyses were carried out using an ESC (unspun) data set comprising 388,389,043 total reads obtained from aggregating two biological replicates, an ESC (spun) data set with 384,462,519 reads, and a sperm data set comprising 389,791,625 reads obtained from aggregating three biological replicate data sets.

Motif Analysis

We first trimmed sequencing adaptors, considering reverse complementarity of the paired-end reads, to recover reads sequenced from short DNA fragments. Then, all reads were aligned to mouse genome mm9 by Bowtie2 with default parameters (end-to-end alignment mode and random selection for multiple mappers). Only concordantly aligned read pairs were retained, and the length of the sequenced DNA fragments was calculated according to their mapped coordinates. Reads representing DNA fragments of mononucleosomal MNase (135–165 bp), subnucleosomal MNase (100–130 bp), and TF-sized footprints (<80 bp) were segregated for several analyses, as indicated. We performed peak calling from TF-sized reads against other two groups (as background inputs) using the Model-based Analysis for ChIP-seq. We further selected the top 1,000 most significant peaks and refined their borders by extending 20 bp both up- and downstream of the peak summits, which are the points with highest read coverage within peaks. We then searched for enriched motifs by using MEME (Multiple EM for Motif Elicitation) among the genomic sequences of the refined peaks. Enriched *de novo* motifs were compared to known motifs using TOMTOM with the motif frequency models from Jaspar.

FISH

Sperm cells were collected by swim-up as previously described, and slides were prepared as described by van der Heijden et al. (2006). Sperm cells placed on coverslips were permeabilized in CSK buffer, 5% triton, and vanadyl ribonucleoside complex for 3 min; then, they were fixed in 4% paraformaldehyde for 10 min and stored in 70% EtOH. Antibodies used for immunohistochemistry were Pan-H4 (Upstate, catalog number 05-858, lot number 27171) and TH2B (Abcam, catalog number ab23913, lot number 825930). Primary antibodies were detected with anti-rabbit 488 (Jackson ImmunoResearch). Slides were incubated with 1:250 primary antibody in 1% BSA, 1× PBS for

1 hr at 37°C and then washed and immunodetected using a 1:500 dilution of conjugated (Alexa Fluor 488 or Alexa Fluor 594, Invitrogen) secondary (anti-goat, -mouse, or -rabbit) antibody in 1× PBS with 1% BSA. HpaII probe was prepared by performing an overnight restriction enzyme digest on purified mouse sperm DNA, followed by gel purification of 50 to 500 bp DNA fragments, ligation of adapters, and PCR amplification in the presence of biotin-dCTP. Mouse CoT-1 DNA (Invitrogen) probe was nick-translated using biotin-11-dUTP (Roche). Two hundred nanograms of HpaII biotinylated probe or 50 ng CoT-1 biotin nick-translated probe was hybridized to fixed and denatured cells.

ACCESSION NUMBERS

Data will be available via the Gene Expression Omnibus database under accession ID GSE58101.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and can be found with this article online at <http://dx.doi.org/10.1016/j.devcel.2014.05.024>.

ACKNOWLEDGMENTS

We thank K. Ahmad and members of the O.J.R. laboratory for insightful discussions and comments on the manuscript. O.J.R. is supported in part by grant HD080224 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), and by the Harold and Leila Mathers Charitable Foundation. T.G.F. is supported in part by grant HD072122 from NICHD and is a Pew Scholar in the Biomedical Sciences. Z.W. is supported by National Science Foundation grant DBI-0850008. B.R.C. is supported by a T32HD007439 training grant. S.J.H. is supported by a T32CA130807 training grant and by a Leukemia and Lymphoma postdoctoral fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Received: March 14, 2014

Revised: May 28, 2014

Accepted: May 30, 2014

Published: July 3, 2014

REFERENCES

- Arpanahi, A., Brinkworth, M., Iles, D., Krawetz, S.A., Paradowska, A., Platts, A.E., Saida, M., Steger, K., Tedder, P., and Miller, D. (2009). Endonuclease-sensitive regions of human spermatozoal chromatin are highly enriched in promoter and CTCF binding sequences. *Genome Res.* 19, 1338–1349.
- Brykczynska, U., Hisano, M., Erkek, S., Ramos, L., Oakeley, E.J., Roloff, T.C., Beisel, C., Schübeler, D., Stadler, M.B., and Peters, A.H. (2010). Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nat. Struct. Mol. Biol.* 17, 679–687.
- Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133, 1106–1117.
- Crick, F.H., and Klug, A. (1975). Kinky helix. *Nature* 255, 530–533.
- Erkek, S., Hisano, M., Liang, C.Y., Gill, M., Murr, R., Dieker, J., Schübeler, D., van der Vlag, J., Stadler, M.B., and Peters, A.H. (2013). Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nat. Struct. Mol. Biol.* 20, 868–875.
- Fazio, T.G., Huff, J.T., and Panning, B. (2008). An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *Cell* 134, 162–174.
- Fu, Y., Sinha, M., Peterson, C.L., and Weng, Z. (2008). The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet.* 4, e1000138.

- Gardiner-Garden, M., Ballesteros, M., Gordon, M., and Tam, P.P. (1998). Histone- and protamine-DNA association: conservation of different patterns within the beta-globin domain in human sperm. *Mol. Cell. Biol.* **18**, 3350–3356.
- Govin, J., Escoffier, E., Rousseaux, S., Kuhn, L., Ferro, M., Thévenon, J., Catena, R., Davidson, I., Garin, J., Khochbin, S., and Caron, C. (2007). Pericentric heterochromatin reprogramming by new histone variants during mouse spermiogenesis. *J. Cell Biol.* **176**, 283–294.
- Hammoud, S.S., Nix, D.A., Zhang, H., Purwar, J., Carrell, D.T., and Cairns, B.R. (2009). Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460**, 473–478.
- Henikoff, S., Henikoff, J.G., Sakai, A., Loeb, G.B., and Ahmad, K. (2009). Genome-wide profiling of salt fractions maps physical properties of chromatin. *Genome Res.* **19**, 460–469.
- Henikoff, J.G., Belsky, J.A., Krassovsky, K., MacAlpine, D.M., and Henikoff, S. (2011). Epigenome characterization at single base-pair resolution. *Proc. Natl. Acad. Sci. USA* **108**, 18318–18323.
- Ihara, M., Meyer-Ficca, M.L., Leu, N.A., Rao, S., Li, F., Gregory, B.D., Zalenskaya, I.A., Schultz, R.M., and Meyer, R.G. (2014). Paternal poly (ADP-ribose) metabolism modulates retention of inheritable sperm histones and early embryonic gene expression. *PLoS Genet.* **10**, e1004317.
- Jelinic, P., Stehle, J.C., and Shaw, P. (2006). The testis-specific factor CTCFL cooperates with the protein methyltransferase PRMT7 in H19 imprinting control region methylation. *PLoS Biol.* **4**, e355.
- Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.* **10**, 161–172.
- Kent, N.A., Adams, S., Moorhouse, A., and Paszkiewicz, K. (2011). Chromatin particle spectrum analysis: a method for comparative chromatin structure analysis using paired-end mode next-generation DNA sequencing. *Nucleic Acids Res.* **39**, e26.
- Kierszenbaum, A.L., and Tres, L.L. (1975). Structural and transcriptional features of the mouse spermatid genome. *J. Cell Biol.* **65**, 258–270.
- Li, Y., Lalancette, C., Miller, D., and Krawetz, S.A. (2008). Characterization of nucleohistone and nucleoprotamine components in the mature human sperm nucleus. *Asian J. Androl.* **10**, 535–541.
- Li, Z., Gadue, P., Chen, K., Jiao, Y., Tuteja, G., Schug, J., Li, W., and Kaestner, K.H. (2012). Foxa2 and H2A.Z mediate nucleosome depletion during embryonic stem cell differentiation. *Cell* **151**, 1608–1616.
- Loukinov, D.I., Pugacheva, E., Vatolin, S., Pack, S.D., Moon, H., Chernukhin, I., Mannan, P., Larsson, E., Kanduri, C., Vostrov, A.A., et al. (2002). BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. *Proc. Natl. Acad. Sci. USA* **99**, 6806–6811.
- Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell* **10**, 105–116.
- Meyer-Ficca, M.L., Lonchar, J.D., Ihara, M., Bader, J.J., and Meyer, R.G. (2013). Alteration of poly(ADP-ribose) metabolism affects murine sperm nuclear architecture by impairing pericentric heterochromatin condensation. *Chromosoma* **122**, 319–335.
- Montellier, E., Boussouar, F., Rousseaux, S., Zhang, K., Buchou, T., Fenaille, F., Shiota, H., Debernardi, A., Héry, P., Curtet, S., et al. (2013). Chromatin-to-nucleoprotamine transition is controlled by the histone H2B variant TH2B. *Genes Dev.* **27**, 1680–1692.
- Nakamura, T., Liu, Y.J., Nakashima, H., Umehara, H., Inoue, K., Matoba, S., Tachibana, M., Ogura, A., Shinkai, Y., and Nakano, T. (2012). PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature* **486**, 415–419.
- Namekawa, S.H., Park, P.J., Zhang, L.F., Shima, J.E., McCarrey, J.R., Griswold, M.D., and Lee, J.T. (2006). Postmeiotic sex chromatin in the male germline of mice. *Curr. Biol.* **16**, 660–667.
- Noll, M., and Kornberg, R.D. (1977). Action of micrococcal nuclease on chromatin and the location of histone H1. *J. Mol. Biol.* **109**, 393–404.
- Ooi, S.L., and Henikoff, S. (2007). Germline histone dynamics and epigenetics. *Curr. Opin. Cell Biol.* **19**, 257–265.
- Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. *Cell* **137**, 1194–1211.
- Pittoggi, C., Renzi, L., Zaccagnini, G., Cimini, D., Degraffi, F., Giordano, R., Magnano, A.R., Lorenzini, R., Lavia, P., and Spadafora, C. (1999). A fraction of mouse sperm chromatin is organized in nucleosomal hypersensitive domains enriched in retroposon DNA. *J. Cell Sci.* **112**, 3537–3548.
- Radman-Livaja, M., and Rando, O.J. (2010). Nucleosome positioning: how is it established, and why does it matter? *Dev. Biol.* **339**, 258–266.
- Saida, M., Iles, D., Elnefati, A., Brinkworth, M., and Miller, D. (2011). Key gene regulatory sequences with distinctive ontological signatures associate with differentially endonuclease-accessible mouse sperm chromatin. *Reproduction* **142**, 73–86.
- Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887–898.
- Sleutels, F., Soochit, W., Bartkuhn, M., Heath, H., Dienstbach, S., Bergmaier, P., Franke, V., Rosa-Garrido, M., van de Nobelen, S., Caesar, L., et al. (2012). The male germ cell gene regulator CTCFL is functionally different from CTCF and binds CTCF-like consensus sites in a nucleosome composition-dependent manner. *Epigenetics Chromatin* **5**, 8.
- Tang, J.B., and Chen, Y.H. (2006). Identification of a tyrosine-phosphorylated CCCTC-binding nuclear factor in capacitated mouse spermatozoa. *Proteomics* **6**, 4800–4807.
- Teif, V.B., Vainshtein, Y., Caudron-Herger, M., Mallm, J.P., Marth, C., Höfer, T., and Rippe, K. (2012). Genome-wide nucleosome positioning during embryonic stem cell development. *Nat. Struct. Mol. Biol.* **19**, 1185–1192.
- Teves, S.S., and Henikoff, S. (2011). Heat shock reduces stalled RNA polymerase II and nucleosome turnover genome-wide. *Genes Dev.* **25**, 2387–2397.
- Valouev, A., Johnson, S.M., Boyd, S.D., Smith, C.L., Fire, A.Z., and Sidow, A. (2011). Determinants of nucleosome organization in primary human cells. *Nature* **474**, 516–520.
- van der Heijden, G.W., Derijck, A.A., Ramos, L., Giele, M., van der Vlag, J., and de Boer, P. (2006). Transmission of modified nucleosomes from the mouse male germline to the zygote and subsequent remodeling of paternal chromatin. *Dev. Biol.* **298**, 458–469.
- Vavouri, T., and Lehner, B. (2011). Chromatin organization in sperm may be the major functional consequence of base composition variation in the human genome. *PLoS Genet.* **7**, e1002036.
- Ward, W.S., and Zalsky, A.O. (1996). The unique, complex organization of the transcriptionally silent sperm chromatin. *Crit. Rev. Eukaryot. Gene Expr.* **6**, 139–147.
- Weiner, A., Hughes, A., Yassour, M., Rando, O.J., and Friedman, N. (2010). High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. *Genome Res.* **20**, 90–100.
- Xi, Y., Yao, J., Chen, R., Li, W., and He, X. (2011). Nucleosome fragility reveals novel functional states of chromatin and poises genes for activation. *Genome Res.* **21**, 718–724.
- Yildirim, O., Li, R., Hung, J.H., Chen, P.B., Dong, X., Ee, L.S., Weng, Z., Rando, O.J., and Fazio, T.G. (2011). Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* **147**, 1498–1510.

Developmental Cell, Volume 30

Supplemental Information

High-Resolution Mapping of Chromatin Packaging in Mouse Embryonic Stem Cells and Sperm

Benjamin R. Carone, Jui-Hung Hung, Sarah J. Hainer, Min-Te Chou, Dawn M. Carone,
Zhiping Weng, Thomas G. Fazzio, and Oliver J. Rando

Supplementary materials

Figures S1-S5

Supplementary figure legends

Figure S1. Genome-scale views of ES and sperm MNase footprints. Related to Figure 3. A-B.

UCSC genome browser tracks are shown for 1-80 and for 135-165 bp footprints, for ES cells and sperm, as indicated. Locus **B** covers the *Hoxd* cluster, revealing sperm nucleosome footprint depletion over *Hoxd* promoters. Note that 1-80 bp footprints are widespread in both cell types. These footprints thus likely represent a combination of products of extensive nuclease digestion of nucleosomes, as well as footprints of specific DNA-binding proteins such as TFs (see **Figures 5-6**). As these abundant short fragments are observed both in ES cells and sperm, they are unlikely to represent the transitional packaging state observed during earlier stages of spermatogenesis (see, eg, (Montellier et al., 2013)). **C-D**. Nucleosome-length MNase footprints in sperm correspond to nucleosomes. Chromatin isolated by anti-histone H3 ChIP was sequenced to modest depth (~1.45 million reads), and H3 signal was analyzed relative to gene density at large scale (**D**) and relative to TSSs at higher resolution (**C**).

Figure S2. MNase-resistant nucleosomes are found at CpG-rich promoters. Related to Figure 4 A.

Data for 150 bp MNase footprints resulting from extensive (50 Unit) overdigestion of sperm chromatin were aligned to all genic TSSs, grouped according to high, intermediate, and low CpG content (HCP, ICP, and LCP, respectively). Data here are from the unfractionated library presented in **Figure 4D**. **B-C**. TSS-aligned 150 bp footprint data are shown for the indicated MNase digestion levels. Here, MNase digestions were centrifuged prior to DNA isolation, and libraries were constructed from supernatant (**B**) or pellet (**C**) material. Both graphs use the same y axis scale, showing that promoter enrichment of MNase footprints observed after extensive MNase digestion is strongest in the pellet fraction.

Figure S3. Short MNase footprints correspond to in vivo TF binding. Related to Figure 5. 1-80 bp

MNase protected fragments were extracted from our ES and sperm datasets. Footprint data here is averaged across all motifs bound in ES cells for the indicated TFs – in vivo binding data was extracted from (Chen et al., 2008; Garber et al., 2012).

Figure S4. Validation of CTCF binding at sites of short MNase footprints in ES cells. Related to

Figure 6. **A**. 9 CTCF motif matches were selected based on occupancy of <80 bp MNase digestion fragments. “Sperm” sites contained <80 bp reads in sperm samples but not ES cells, “ESC” sites were ES-specific, and “Both” contained <80 bp reads in both sperm and ES libraries. As shown, significant anti-CTCF ChIP signal in sonicated mESC chromatin is only evident at sites in predicted to contain CTCF based on the presence of <80bp MNase DNA fragments in ES cells. **B-C**. V plots, generated as in **Figures 5B-C**, were generated for CTCF motifs for EGFP and CTCF knockdown ES cells, as indicated. **D**. Western blots of CTCF and Histone H3 in testis and sperm lysates. CTCF and Histone H3 protein are both evident in testis and sperm samples. Anti-CTCFL westerns showed no apparent signal in sperm (not shown).

Figure S5. <80 bp MNase digestion products are enriched over CTCF but not CTCFL sites in

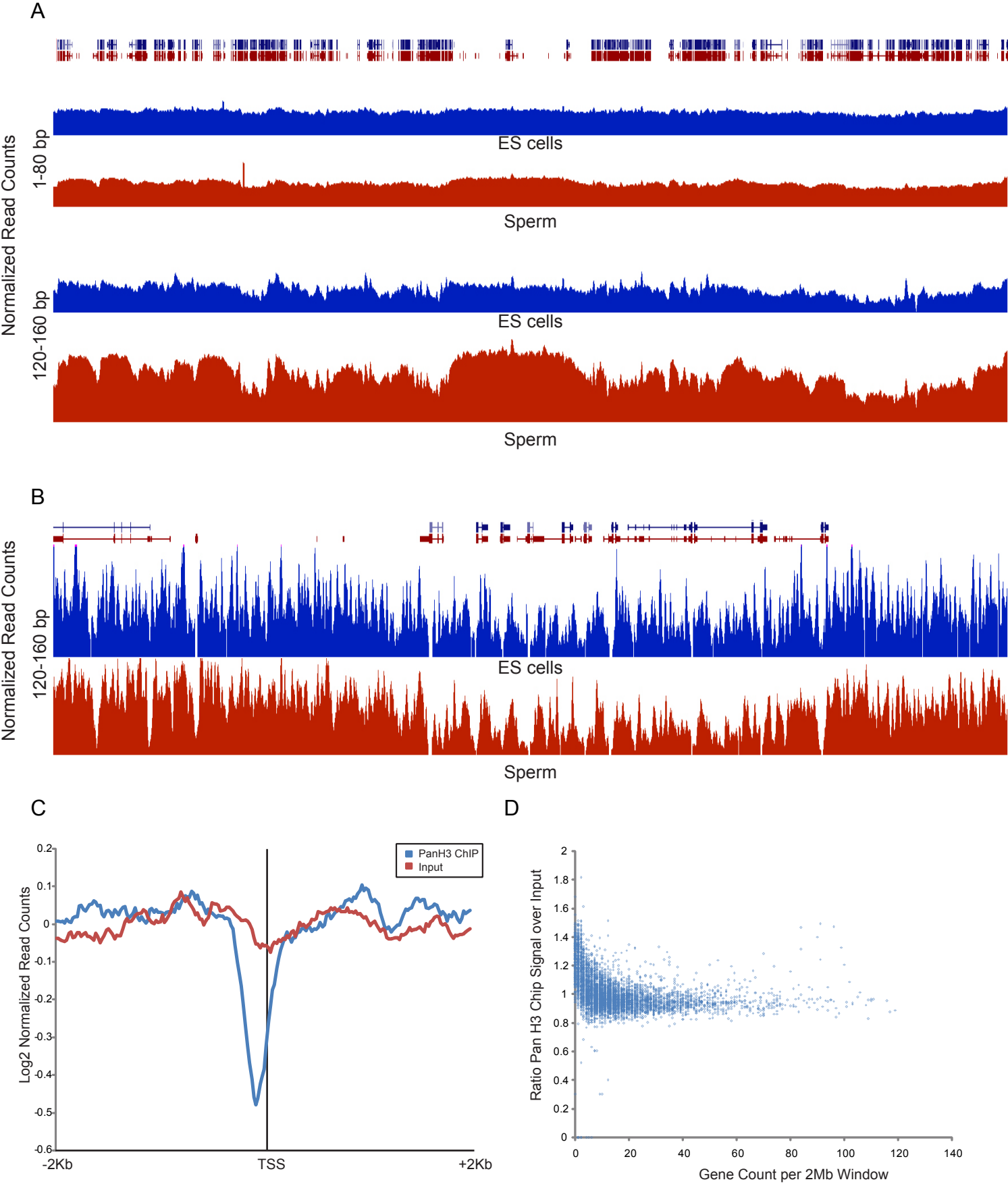
sperm. Related to **Figure 6**. **A**. Chip signal from CTCF or CTCFL pulldown (Sleutels et al.) aligned over all predicted CTCF or CTCFL binding sites identified in Sleutels *et al.* While CTCF ChIP signal dominates at both binding sites, only at CTCFL binding sites is CTCFL ChIP signal detected. **B**. Designation of CTCFL or CTCF mutually exclusive binding sites from CTCF/CTCFL ratio, approximately 1500 sites were

used for both classes. **C.** Enrichment of <80 bp MNase fragments is evident over both CTCF and CTCFL sites in mESCs, while only CTCF-specific sites in sperm exhibit strong <80 bp occupancy.

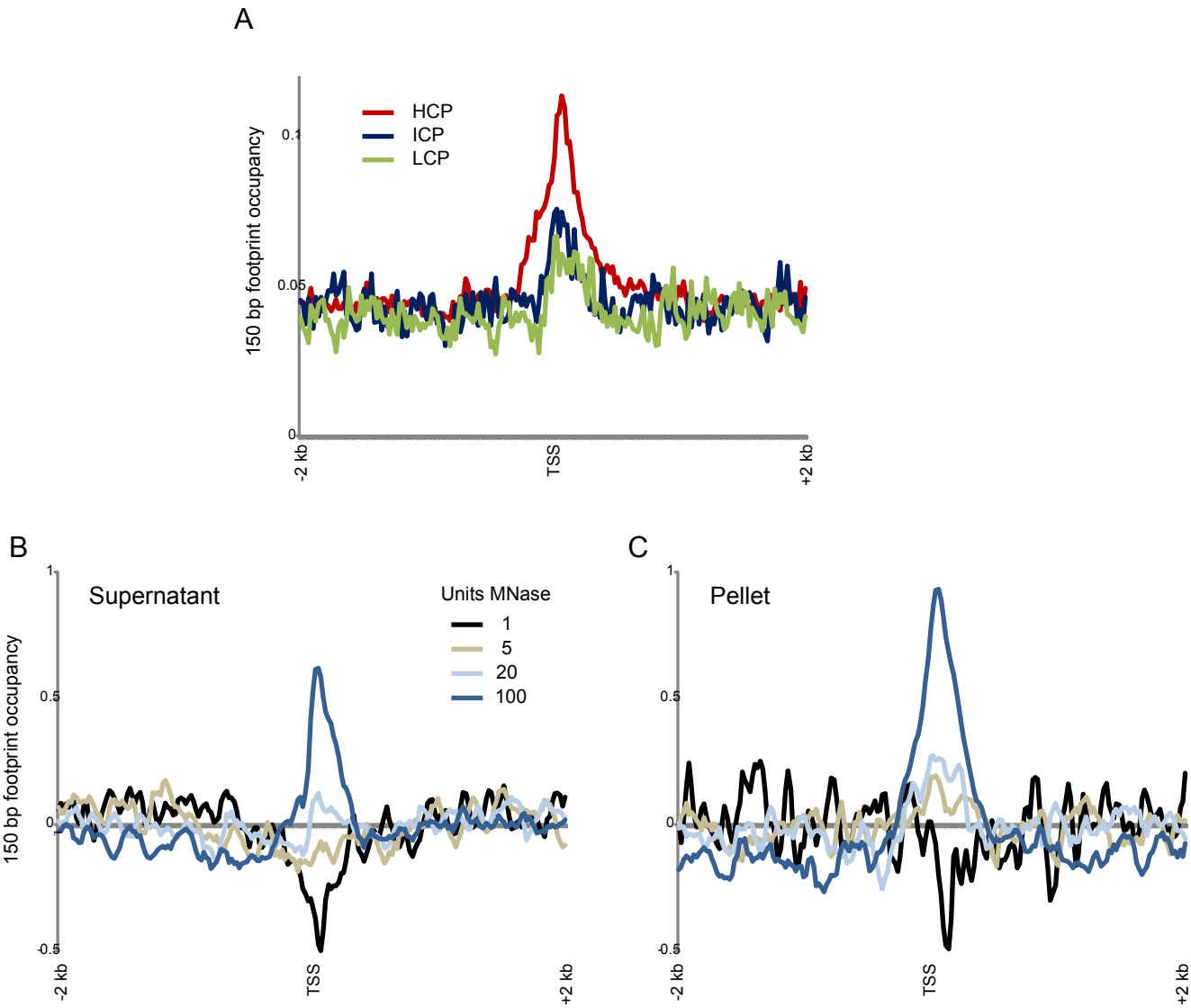
References

- Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., *et al.* (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133, 1106-1117.
- Garber, M., Yosef, N., Goren, A., Raychowdhury, R., Thielke, A., Guttman, M., Robinson, J., Minie, B., Chevrier, N., Itzhaki, Z., *et al.* (2012). A High-Throughput Chromatin Immunoprecipitation Approach Reveals Principles of Dynamic Gene Regulation in Mammals. *Molecular cell*.
- Montellier, E., Boussouar, F., Rousseaux, S., Zhang, K., Buchou, T., Fenaille, F., Shiota, H., Debernardi, A., Hery, P., Curtet, S., *et al.* (2013). Chromatin-to-nucleoprotamine transition is controlled by the histone H2B variant TH2B. *Genes & development* 27, 1680-1692.
- Sleutels, F., Soochit, W., Bartkuhn, M., Heath, H., Dienstbach, S., Bergmaier, P., Franke, V., Rosa-Garrido, M., van de Nobelen, S., Caesar, L., *et al.* (2012). The male germ cell gene regulator CTCFL is functionally different from CTCF and binds CTCF-like consensus sites in a nucleosome composition-dependent manner. *Epigenetics & chromatin* 5, 8.

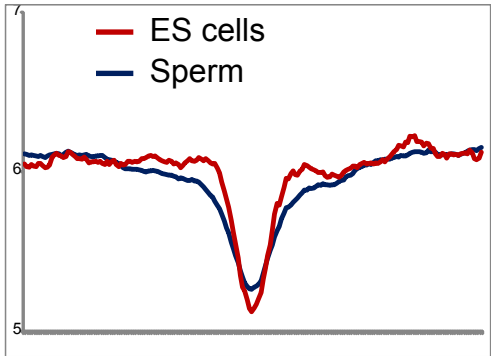
Supplemental Fig. S1



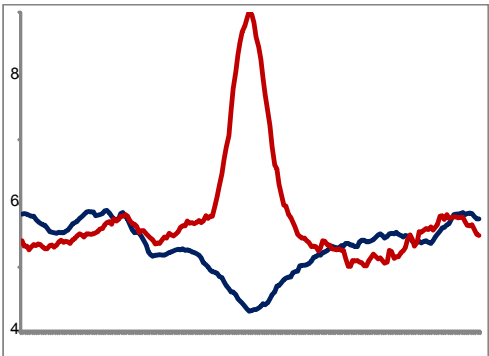
Supplemental Fig. S2



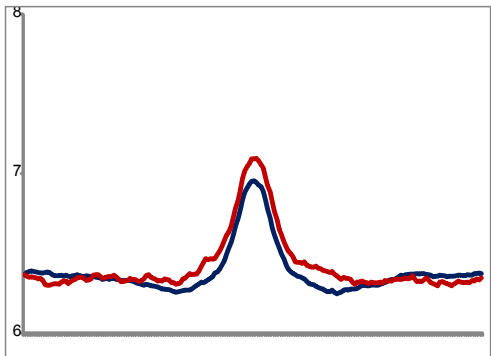
Supplemental Fig. S3



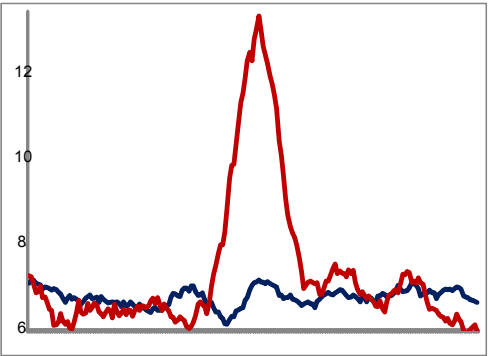
STAT1



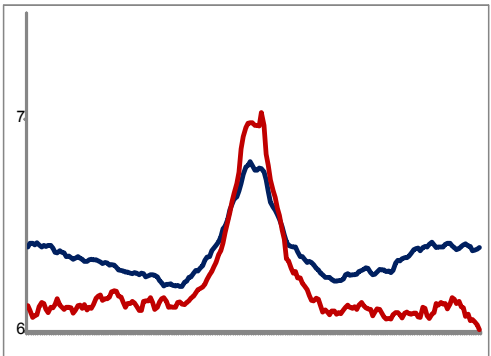
KLF4



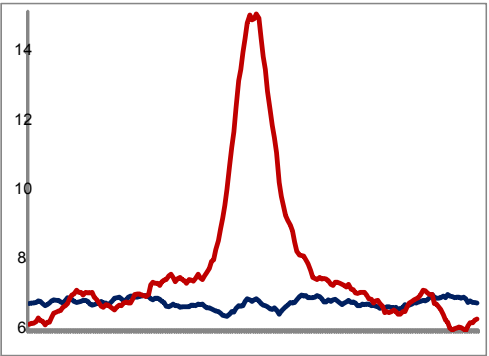
RelA



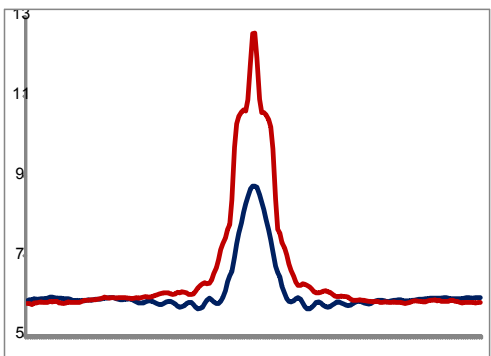
Nanog



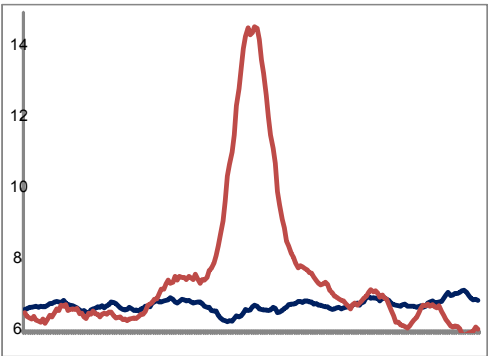
STAT3



OCT4

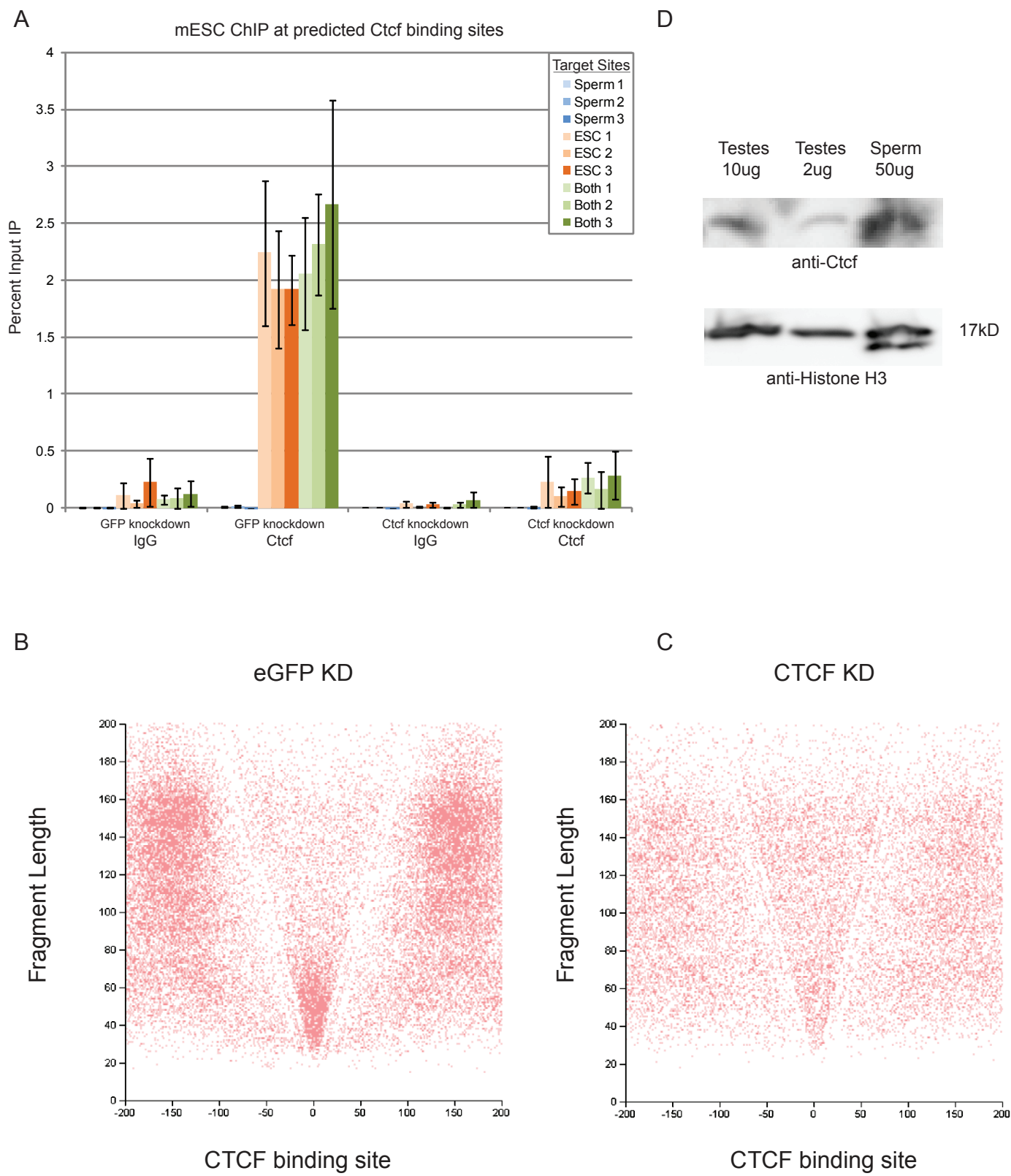


CTCF



SOX2

Supplemental Fig. S4



Supplemental Fig. S5

