RESEARCH ARTICLE

Nucleosome Organization in Human Embryonic Stem Cells

Puya G. Yazdi[1,2,3☯¤], **Brian A. Pedersen**[1,2,3☯], **Jared F. Taylor**[1,2,3], **Omar S. Khattab**[1,2], **Yu-Han Chen**[1,2,3], **Yumay Chen**[1,2,3], **Steven E. Jacobsen**[6,7,8], **Ping H. Wang**[1,2,3,4,5]*

1 UC Irvine Diabetes Center, University of California Irvine, Irvine, California, United States of America, **2** Sue and Bill Gross Stem Cell Research Center, University of California Irvine, Irvine, California, United States of America, **3** Department of Medicine, University of California Irvine, Irvine, California, United States of America, **4** Department of Biological Chemistry, University of California Irvine, Irvine, California, United States of America, **5** Department of Physiology & Biophysics, University of California Irvine, Irvine, California, United States of America, **6** Department of Molecular, Cell and Developmental Biology, University of California Los Angeles, Los Angeles, California, United States of America, **7** Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California Los Angeles, Los Angeles, California, United States of America, **8** Howard Hughes Medical Institute, University of California Los Angeles, Los Angeles, California, United States of America

☯ These authors contributed equally to this work.
¤ Current Address: Systomic Health LLC, Los Angeles, California, United States of America
* phwang@uci.edu

Abstract

The fundamental repeating unit of eukaryotic chromatin is the nucleosome. Besides being involved in packaging DNA, nucleosome organization plays an important role in transcriptional regulation and cellular identity. Currently, there is much debate about the major determinants of the nucleosome architecture of a genome and its significance with little being known about its role in stem cells. To address these questions, we performed ultra-deep sequencing of nucleosomal DNA in two human embryonic stem cell lines and integrated our data with numerous epigenomic maps. Our analyses have revealed that the genome is a determinant of nucleosome organization with transcriptionally inactive regions characterized by a "ground state" of nucleosome profiles driven by underlying DNA sequences. DNA sequence preferences are associated with heterogeneous chromatin organization around transcription start sites. Transcription, histone modifications, and DNA methylation alter this "ground state" by having distinct effects on both nucleosome positioning and occupancy. As the transcriptional rate increases, nucleosomes become better positioned. Exons transcribed and included in the final spliced mRNA have distinct nucleosome profiles in comparison to exons not included at exon-exon junctions. Genes marked by the active modification H3K4m3 are characterized by lower nucleosome occupancy before the transcription start site compared to genes marked by the inactive modification H3K27m3, while bivalent domains, genes associated with both marks, lie exactly in the middle. Combinatorial patterns of epigenetic marks (chromatin states) are associated with unique nucleosome profiles. Nucleosome organization varies around transcription factor binding in enhancers versus promoters. DNA methylation is associated with increasing nucleosome occupancy and different types of methylations have distinct location preferences within the nucleosome core particle. Finally, computational analysis of nucleosome organization alone is sufficient

to elucidate much of the circuitry of pluripotency. Our results, suggest that nucleosome organization is associated with numerous genomic and epigenomic processes and can be used to elucidate cellular identity.

Introduction

Pluripotent stem cells hold great promise in regenerative medicine due to their ability to differentiate into all three germ layers: endoderm, mesoderm, and ectoderm. Human pluripotent stem cells can be divided into embryonic stem cells (hESC), which are derived from the inner cell mass of a blastocyst, and induced pluripotent stem cells (iPSC), which are generated or "reprogrammed" directly from somatic cells[1, 2]. To fully develop the possible therapeutic potential of stem cells, considerable research has been undertaken to study the role epigenetic modifications play in maintaining pluripotency and inducing differentiation. Additionally, recent work has demonstrated that while somatic and pluripotent cells share many similar epigenomic characteristics, there are unique features in the epigenome of embryonic stem cells[3–11]. While much of this work has focused on DNA methylation and chromatin modifications, epigenomic analysis of the primary unit of chromatin, the nucleosome, is scarce.

In eukaryotes, DNA is packaged into chromatin whose fundamental repeating unit is the nucleosome. The nucleosome is comprised of two copies of each of the core histones (H2A, H2B, H3, and H4) wrapped by 147 base pairs (bp) of DNA, with the symmetrical center being called the dyad[12]. Besides being involved in packaging DNA, nucleosome positioning (the genomic location of nucleosomes), nucleosome occupancy (how enriched a genomic location is for nucleosomes), and epigenetic modifications (post-translational modifications of histone proteins and DNA methylation) are thought to play a role in development, transcriptional regulation, cellular identity, evolution, and human disease[13–21]. Analyses in model organisms and humans have revealed that the nucleosome organization of a genome is affected by such diverse factors as underlying DNA sequences, nucleosome remodelers, protein binding, and the transcriptional machinery[13–18, 22–31]. Currently there is considerable debate about the roles and extent these factors play, especially in humans compared to yeast[32–36]. Furthermore, to the best of our knowledge, no one has generated genome-wide maps of nucleosomes in hESC and analyzed its potential role in pluripotency. To begin addressing these questions, we paired-end sequenced Micrococcal Nuclease (MNase) digested DNA from H1 and H9 human embryonic stem cells (hESC), yielding 180x and 70x depth of coverage of the human genome, respectively. A nucleosome occupancy score (NOS) map at single bp resolution without smoothing was calculated and used to call nucleosomes (Methods)[37]. The same processing was performed on ten other non-hESC datasets including one *in vitro* (IV) dataset, derived by reconstituting recombinant histones with genomic DNA from human granulocytes as a measure of the purely sequence driven component of nucleosome organization[17]. Additionally, nucleosome data was analyzed against a diverse set of epigenomic and genomic features [20, 38–40]. Finally, nucleosome architecture alone was used to predict transcription factor binding sites.

Results

Nucleosome map generation

H1 and H9 human embryonic stem cells were utilized to generate paired-end MNase-Seq data. After MNase digestion, nucleosomal DNA was visualized on a 2% agarose gel to assess for

laddering of mono-, di- and tri-nucleosomal fragments (S1A Fig). Densitometry of the nucleso-mal DNA from an image captured with a UV light box from one representative gel demonstrated that the band corresponding to mononucleosomal DNA (~150bp) was 71% of total DNA (S1A Fig). As can be seen by the DarkReader images and our methods, we titrated various amounts of MNase digestion, corresponding to 70–90% mononucleosomal DNA since this is considered the ideal amount of digestion[14–17]. As UV light can cause cross-linking of DNA and diminish the quality of DNA, the DNA used to make the next-generation sequencing libraries was visualized with a DarkReader (blue LED transilluminator) (S1B Fig). Two biological replicates for both H1 and H9 were performed, each consisting of six technical replicates for H1 and two technical replicates for H9. The two biological replicates were from two separate next-generation DNA sequencing runs (R51 and R54). 2,586,825,651 and 869,318,927 raw paired-end reads from H1 and H9 cells were sequenced, respectively. This corresponds to an average depth of coverage of approximately 180x for the H1 cell line and approximately 70x for the H9 cell line. The total number of raw reads and the alignment data is shown in S1 Table.

We then compared all technical and biological replicates. Specifically, we performed genome-wide Pearson correlation coefficients (PCC) of the BAM files from each replicate. Please see S2 and S3 Tables, as well as S2 and S3 Figs for the data presented as a table and as a heatmap, respectively. This analysis compares all individual H1 replicates to one another, between sequencing runs, to the pooled H1 datasets, and includes a comparison of the sequencing runs. The same was also done for the H9 replicates. Additionally, the pooled H1 and H9 datasets from both sequencing runs were compared. The minimum PCC between all individual H1 and H9 replicates is 0.892 and 0.982, respectively. The average PCC for this comparison for all H1 replicates was 0.969 and for all H9 replicates was 0.989. Finally, the pooled H1 and pooled H9 datasets were directly compared and had a PCC of 0.961. Based on the high similarity between the H1 replicates, all H1 replicates were combined for downstream analysis according to ENCODE guidelines[38]. As this was also true for H9, they were also combined. The DANPOS algorithm has previously been shown to outperform other nucleosome positioning software and was thus utilized to process all sequencing datasets[37]. After processing the datasets through the DANPOS algorithm, a distribution of nucleosomal fragment sizes for each dataset were compared (S4 Fig). The average fragment size for all datasets was 153.7 bp. The fragment sizes for the paired-end datasets (H1, H9, GM18507, GM18508, GM18516, GM18522, GM19193, GM19238, GM19239) are indeed consistent with mononucleosomal DNA and demonstrate a typical size distribution.

Sequence driven nucleosome organization

We then turned our attention to the role of DNA sequences in nucleosome organization. While considerable work has been done on these features in model organisms, we sought to ascertain if these features are observed in H1 and H9 hESCs. Nucleosomes are enriched for G/Cs and depleted for A/Ts (S5A and S6A Figs). Additionally, AA/TT dinucleotides show ~10 bp periodicities (by fast Fourier transforms (FFT)) confirming that sequence preferences are conserved across eukaryotes (Fig 1A and S6B–S6D Fig)[13, 22, 23]. Interestingly, all other dinucleotides also demonstrated small ~10 bp periodicities (S5B–S5D and S6B–S6D Figs). NOS maps for all 12 datasets were analyzed against 13,912 high confidence protein coding gene coordinates with unique transcription start sites (TSS) (Methods). IV nucleosomes failed to produce well-positioned arrays, a finding confirmed by genome-wide FFT, in line with recent work that shows this is a role of nucleosome remodelers (Fig 1B and 1C and S7 Fig)[18, 26, 27]. Overall, we can conclude that sequence driven components of nucleosome architecture are independent of cell type.
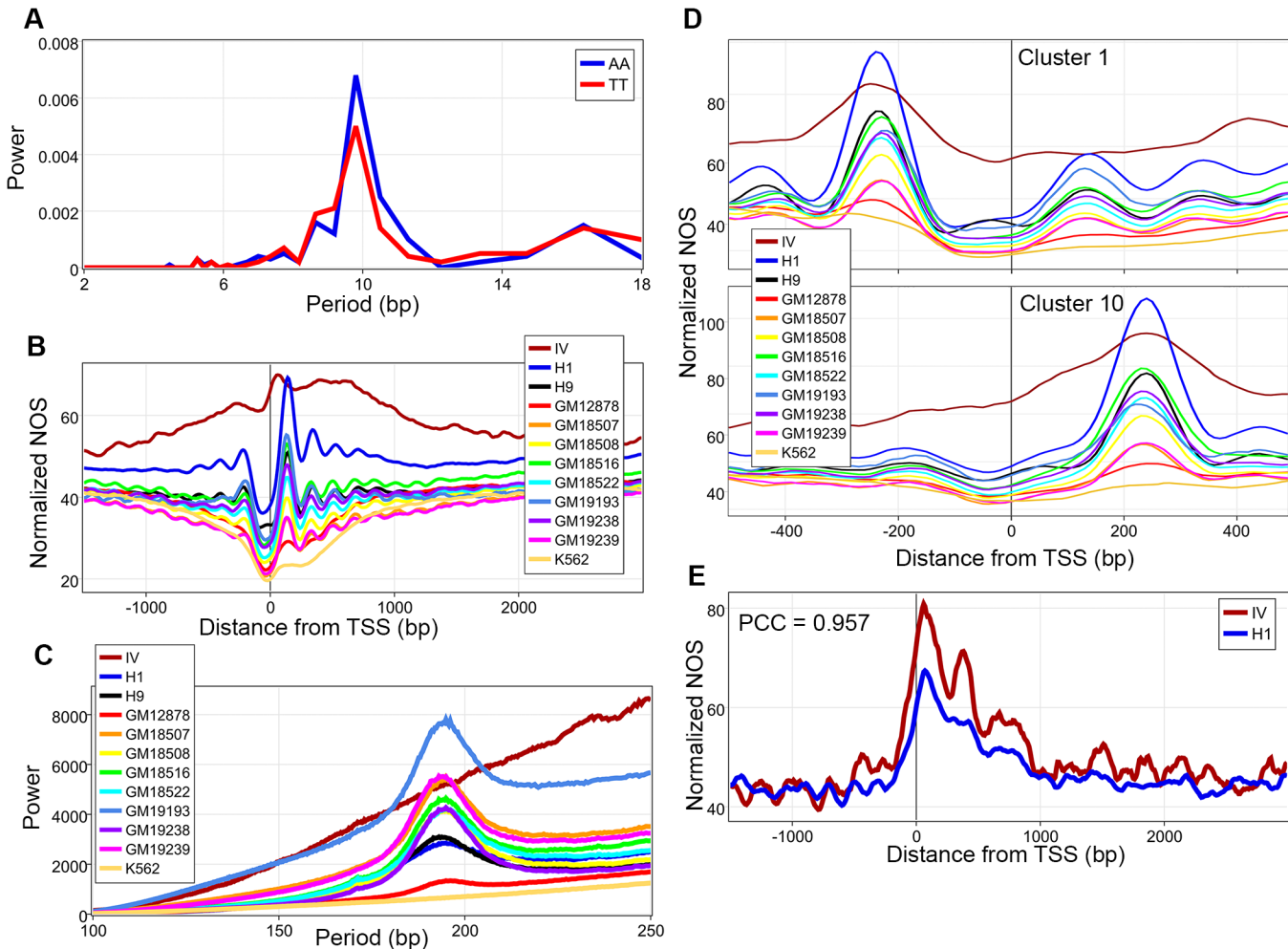
Fig 1. Sequence driven nucleosome organization. A, Fast Fourier transforms (FFT) of AA and TT dinucleotide frequencies through the nucleosome core particle were calculated revealing ~10 bp periodicities. **B**, Normalized nucleosome occupancy scores (NOS) from 12 datasets averaged over 13,912 unique transcription start sites (TSS). **C**, Genome-wide FFT of all 12 datasets showing lack of arrayed nucleosomes in the *in vitro* (IV) dataset. **D**, Normalized NOS for all 12 datasets based on *k*-means clustering of the H1 data, demonstrating that the location of the major peak around TSS is determined by underlying DNA sequences, see S9 Fig for remaining clusters. **E**, IV and H1 NOS were calculated against inactive TSS, the two datasets were highly correlated with a Pearson's correlation coefficient (PCC) of 0.957.

doi:10.1371/journal.pone.0136314.g001

Recently, clustering analysis has shown that chromatin architecture around TSS is heterogeneous[28]. *K*-means clustering of the H1 NOS around TSS revealed 10 clusters that have either a well-positioned downstream or upstream nucleosome, corroborating previous analysis (S8 Fig). Intrigued by what was driving their location, we hypothesized that it could be underlying DNA sequences. To test this, all datasets were analyzed against the coordinates for the 10 different clusters derived from H1. In all 10 clusters, the location of the predominant peak for nucleosome occupancy in each cell line was similar(Fig 1D and S9 Fig). To better visualize this finding, we plotted the location of maximum NOS for all cell lines for all 10 clusters(S10 Fig). The location of the points in this plot are representative of the location of the dyads for the nucleosomes with the highest occupancy within the entire cluster for a given cell line. Based on the clustering analysis, we hypothesized that DNA sequences were a primary driver of nucleosome positions in the absence of transcription. Specifically, the contribution of the underlying

DNA sequence should be greatest in transcriptionally silent regions. Furthermore, nucleosomes are strand independent while transcription is strand specific, and a large portion of the genome is transcribed, regardless of its functional significance[38]. Hence, to accurately determine transcriptionally silent TSS we created a total RNA signal map (RNA-Signal) at single bp resolution by adding the signal from both strands[38]. Based on this map, silent genes were defined and analyzed for both the H1 and IV NOS revealing that the two were highly correlated (PCC = 0.957, Fig 1E). By accounting for known H1 structural variants, the IV and H1 datasets show a genome-wide PCC of 0.695. This suggests that underlying DNA sequences around TSS are highly correlated with nucleosome organization and could create a "ground state" of nucleosome architecture in such regions. Additionally, our genome-wide mononucleotide frequencies and FFT analyses of dinucleotide frequencies within the nucleosome core particle, in conjunction with the nucleosome occupancy correlations between the IV and H1 datasets, implicate underlying DNA sequences play a role in determining nucleosome organization.

Epigenetic regulation of nucleosomes

Next, we sought to address how hESC specific transcription, transcription factor binding, and histone post-translational modifications can alter this "ground state"[38]. We used our RNA-Signal to break up our gene list into quartiles based on total RNA expression and analyzed H1 NOS map against these coordinates and divided the signal by the IV NOS for that gene to accurately quantify how transcription changes the "ground state"[38]. As the transcriptional rate increases the nucleosome depleted region (NDR) becomes less occupied, the +1 nucleosome becomes better positioned with an increased peak, and the nucleosomes become better arrayed, demonstrating how the transcriptional machinery, most likely along with remodelers, work to space nucleosomes and alter the compaction of the chromatin (Fig 2A)[18]. Since the IV dataset might be biased due to being generated with very low assembly degrees (one nucleosome per 850 bp) and on rather short DNA fragments, we repeated this analysis without normalization against the IV dataset. Though less easily discernible, these same findings were found in this analysis (data not shown). We then computed NOS around exons included and excluded in exon-exon splice junctions and found that nucleosomes cover much more of the junction in excluded exons, hinting at their possible role in alternative splicing (Fig 2B). We also find that nucleosome architecture at transcription factor binding sites is only arrayed at active enhancers and not at active promoters (Fig 2C). Coordinates for enhancers and promoter were from the ENCODE dataset[38]. Active sites were defined as those for which the DNase-Seq signal was high, a ChiP-Seq peak was called for a transcription factor, and NOS was low in H1 cells. It must be stated that transcription factor binding sites at active promoters are not necessarily in close proximity to its associated TSS. In fact, the median distance between the transcription factor binding site at active promoters and its associated TSS is 1686 bp for the H1 dataset. We then looked at how histone post-translational modifications can affect nucleosomes. Genes marked by the inactive modification H3K27me3 in their promoters have a higher overall NOS and are less depleted at the NDR compared to genes marked by the active modification H3K4me3 while bivalent genes lie in the middle of the two antagonistic marks (Fig 2D)[41]. Additionally, the location of the +1 nucleosome shifted further downstream, 7 bp from inactive to bivalent and 10 bp from bivalent to active. The average NOS for nucleosomes marked by one of ten histone marks was calculated and showed statistically significant differences for all comparisons (Kruskall-Wallis followed by Mann-Whitney-Wilcoxon with Bonferroni correction, Fig 2E and S5 Table). Finally, NOS was plotted against 15 predicted chromatin state start sites (S11 Fig for definitions) and show that combinatorial patterns of epigenetic marks are correlated with unique and specific nucleosome architectures (Fig 2F and S11 Fig)[20]. Chromatin
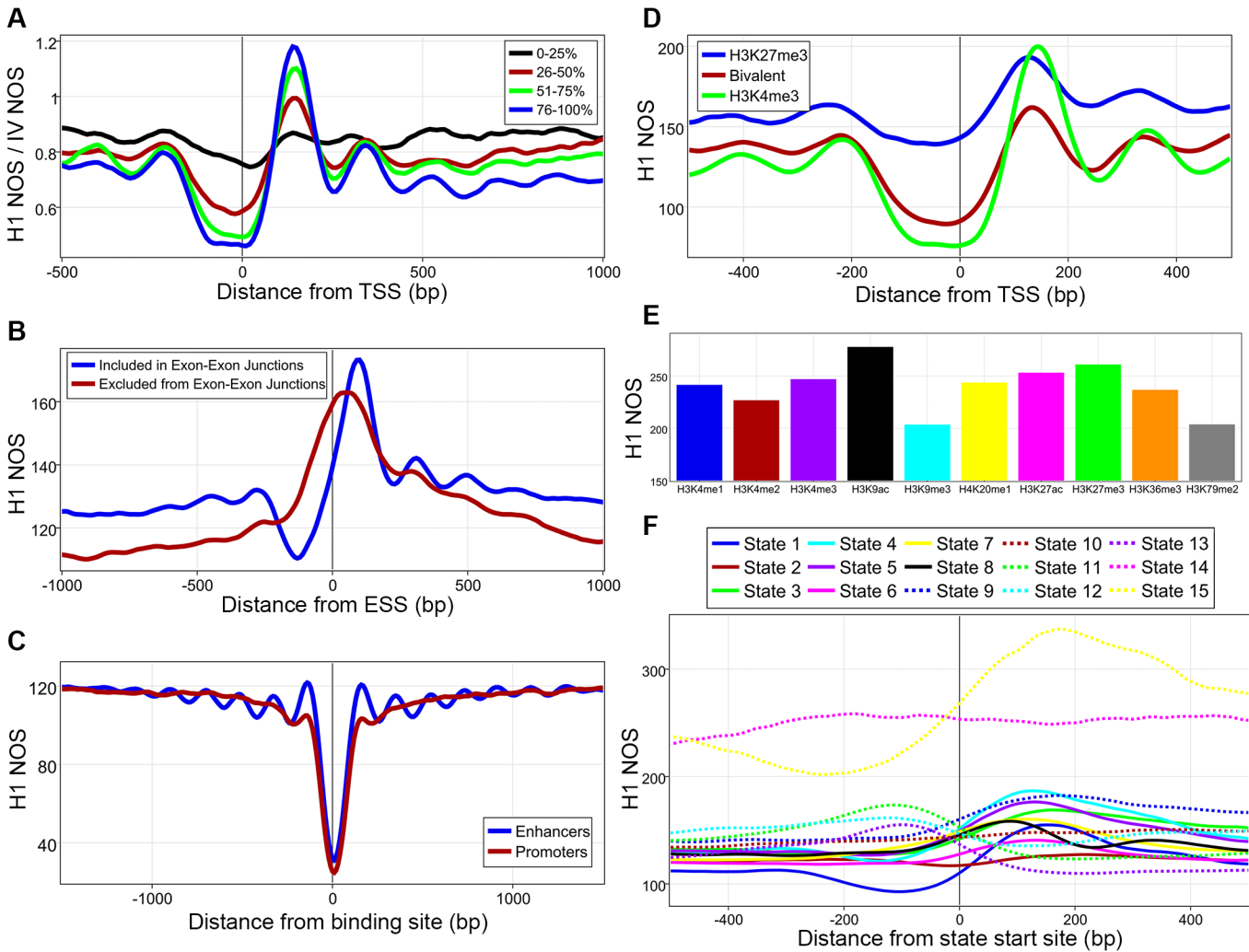
Fig 2. Epigenetic regulation of nucleosomes. A, 13,912 genes were divided into quartiles based on total RNA-Expression from both strands with 0% being lowest total expression and 100% highest expression. B, Nucleosome occupancy score (NOS) plotted against exon start sites (ESS) from exons included in exon-exon junctions and those excluded. C, H1 NOS averaged around transcription factor binding sites at active promoters and active enhancers. D, NOS plotted for genes marked by H3K4me3, H3K27me3, or both (bivalent) in their promoters. E, Average NOS for whole nucleosomes found with one of 10 histone-modifications, all comparisons were statistically significant. F, H1 NOS plotted against 15 different chromatin state start sites, see S11 Fig for definitions.

doi:10.1371/journal.pone.0136314.g002

states are defined computationally by a collection of epigenetic marks. As such, certain states are probably more readily defined than others. Additionally, as these locations are defined by a start site of a region of DNA and not by a transcription factor which binds to the middle of this region, it is impossible to directly compare these chromatin state figures with traditional figures of nucleosomes around promoters, enhancers, and TSS which are generated around the middle of a region. That being said, we did find it interesting that the insulators are associated with well arrayed nucleosomes regardless of these limitations. Overall, we can surmise that active marks and activation of the transcriptional machinery is associated with lower nucleosome occupancy and better-arrayed nucleosomes.
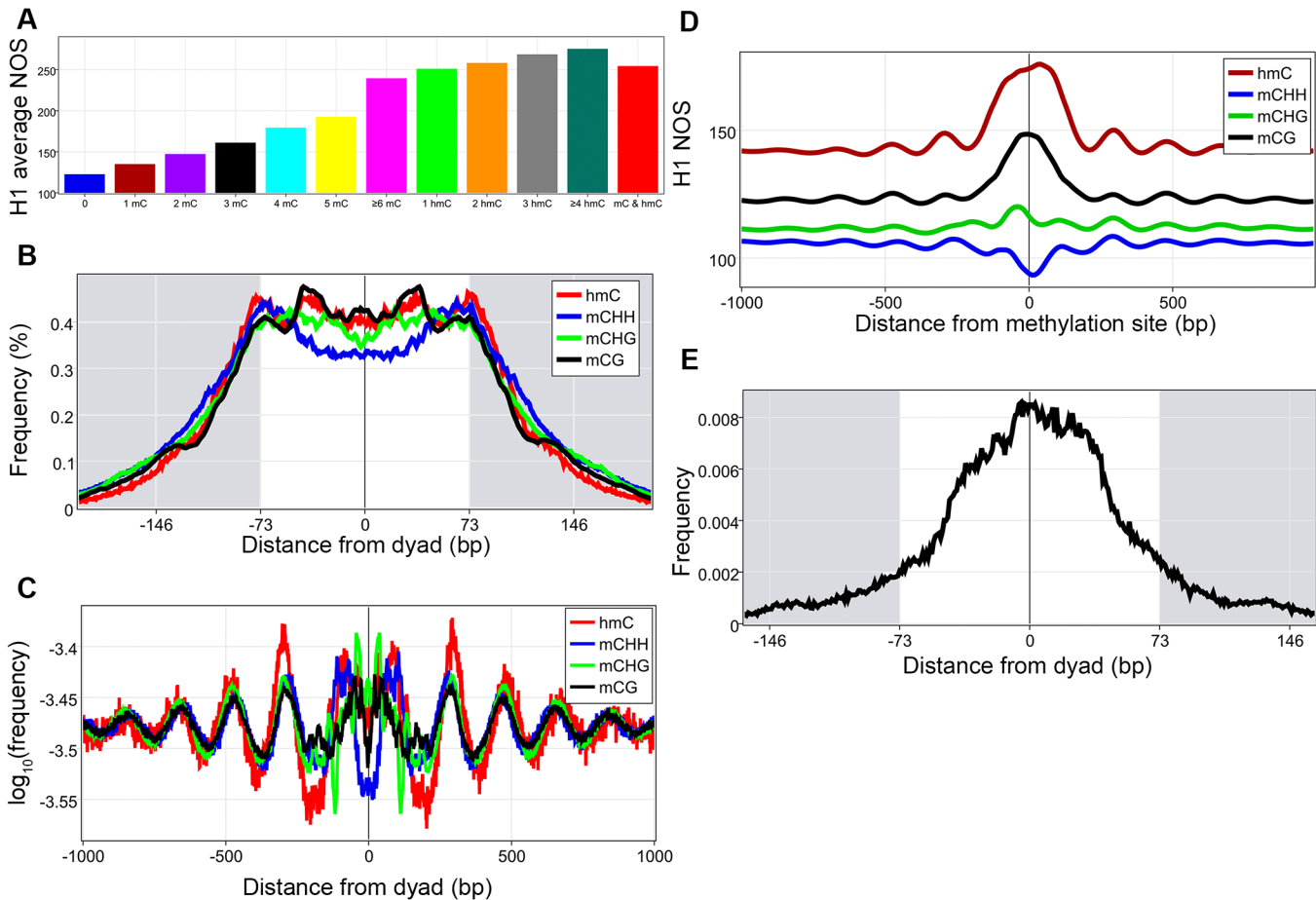
Fig 3. DNA methylation and nucleosomes. **A**, Nucleosomes were divided based on the type and number of methylations found in them and the average nucleosome occupancy score (NOS) was calculated for each group, all comparisons were statistically significant. **B**, The distance from the dyad to the closest methylation was calculated for all mCG, mCHH, mCHG, and hmC separately and converted to frequency percentages. **C**, Plot of all DNA methylations around nucleosome dyads. **D**, NOS were calculated around all mCG, mCHG, mCHH, and hmC sites. **E**, Plot of mCG frequency to closest dyad for mCGs associated with NOS above 190, demonstrating their enrichment toward the dyad.

doi:10.1371/journal.pone.0136314.g003

DNA methylation and nucleosomes

We then addressed how DNA methylation affects nucleosome occupancy[39, 40, 42]. Increasing methylated cytosines found in a nucleosome is associated with an increased average nucleosome occupancy (statistically significant by a Kruskall-Wallis followed by Mann-Whitney-Wilcoxon with Bonferroni correction, Fig 3A and S5 Table). Studies in plants have shown that methylated cytosines are enriched within nucleosomes and display ~10 bp periodicity[43]. To investigate these findings in humans, the distance from each of the four types of methylations (three types of 5-methylcytosine (mCG, mCHG, mCHH, where H = A,T,or C), and 5-hydroxy-methylcytosine (hmC)) to the nearest dyad was plotted, revealing that the four types of methyl-ations have distinct location preferences within the nucleosome (Fig 3B). Interestingly, mCG is enriched at around +/- 40 bp and around the dyad, the three locations that have the strongest DNA nucleosome binding, providing a possible mechanism whereby DNA methylations can increase nucleosome occupancy. FFTs of the four different methylations within the nucleo-some core revealed the periodicity of the signal (S12 Fig)[44]. All methylations against dyads were plotted revealing that on a genome-wide level, methylations are found in a periodic

pattern around nucleosome dyads (Fig 3C). We hypothesized that this periodic methylation pattern might then be associated with periodic nucleosomes and by plotting H1 NOS around methylation sites we demonstrate this with the caveat that mCHHs are associated with a decreased NOS (Fig 3D). Finally, we became intrigued by the possibility that mCGs associated with higher NOS (those above the average peak signal of 190) might have a greater enrichment closer to the dyad thereby increasing the DNA-histone interaction, which our results demonstrate (Fig 3E). This brings about the intriguing possibility that DNA methylation can deactivate genes by creating tightly bound nucleosomes that are an impediment to transcriptional machinery.

Nucleosome architecture and the circuitry of pluripotency

We then turned our attention to the possibility that genome-wide nucleosome maps could be used to deduce the circuitry of transcription factors driving the cell state. Our data along with previous work has demonstrated that transcription factors turn on genes by binding to enhancers and promoters and displacing nucleosomes in the process. This process creates well-defined nucleosome architectures: a missing nucleosome surrounded by two bound nucleosomes that are relatively well-positioned. Hence, we hypothesized that by scanning our nucleosome map for these patterns and then integrating the resulting DNA sequences with motif discovery tools we might be able to ascertain some of the transcription factors that drive the circuitry of pluripotency. Our computational approach was able to predict that Oct4, Sox2, KLF4, and Nanog, classic transcription factors used for reprogramming and believed to be driving the circuitry of pluripotency, are active in our cell line based on nucleosome analysis alone (Fig 4, S13 Fig)[1, 45–47]. Based on this data, it seems plausible that nucleosome analysis could be used as a first step in reprogramming or transdifferentiating different cell types by helping generate a list of active transcription factors driving that cell's circuitry.

Discussion

By integrating our H1, H9, and published somatic cell *in vivo* data with *in vitro* data, we set out to determine how and in what ways underlying DNA sequences are associated with nucleosome organization and if these patterns were similar in pluripotent and somatic cells. The IV dataset was created by reconstituting recombinant histones with DNA from human granulocytes in ~1:1 molar ratio and hence nucleosome occupancy variation across the genome is being driven by underlying DNA sequence preferences alone[17]. Additionally, since ~10 bp periodicity cannot be easily discerned, we followed previously establish methods and utilized a FFT to determine the periodicity, if any[17, 43]. A Fourier transform is a mathematical method, with many different applications, that converts a signal in space into a combination of pure frequencies. As such, FFTs were performed for each dinucleotide to more precisely determine if a periodicity (1/frequency) existed, and if so what it is for each dinucleotide within the nucleosome core particle (Fig 1A, S5D and S6D Figs). Our data corroborates studies in yeast and a recent work in humans that has shown that nucleosomal DNA demonstrates an AA/TT ~10 bp periodicities and is enriched for G/C content (Fig 1A, S5A, S6A and S6D Figs)[13, 22, 23]. Furthermore, we can conclude that on a global level, nucleosome architecture is similar in both somatic and pluripotent cells. This is evidenced by the similarity of the nucleosome architecture around the TSS in all of the datasets (Fig 1D and 1E, S9 and S10 Figs). This is further corroborated by the genome-wide PCC of 0.695 between NOS maps for H1 and IV datasets. These findings suggest that nucleosome organization could be driven by underlying DNA sequences in transcriptionally silent regions(though we cannot rule out that the underlying molecular biology could have biased the correlations), which can lead us to hypothesize that

Gene	Motif	q-value (Benjamini)
Oct4		0.0071
Klf-4		0.0232
Nanog		0.0573
Sox2		0.0799

Fig 4. **Computationally derived active enhancer sites.** HOMER motifs with their associated gene at enhancer binding site locations, as defined by surrounding nucleosome positioning (Methods). Included in the table are genes that maintain the pluripotent state.

doi:10.1371/journal.pone.0136314.g004

the genome drives a primary organization of nucleosome architecture, which the epigenome and transcription can alter in a cell-specific manner.

Taking advantage of the richness of the available ENCODE data, we integrated our H1 data with numerous epigenomic maps[38]. First, we demonstrated how spacing of nucleosomes and ordered arrays around TSS is highly correlated with the transcriptional rate (Fig 2A). Since previous work has shown that nucleosome remodelers are involved in spacing nucleosomes, it appears that remodelers and the transcriptional machinery work together to create the classic nucleosome organization[18, 26, 27]. It is also tempting to speculate that perhaps histone post-translational modifications can be read by nucleosome remodelers and this would allow for a fine-tuning of directionality and needed spacing for transcription. The variable nucleosome architecture around exons falls in line with recent work that has shown that chromatin plays a role in alternative splicing (Fig 2B)[48–50]. Interestingly, transcription factor binding at enhancers and promoters created different chromatin architectures (Fig 2C). Since the classic arrayed pattern was only observed at enhancers, it is tantalizing to hypothesize that perhaps nucleosome organization around active enhancers is involved in three-dimensional structural changes and possibly DNA looping. Both nucleosome positioning and occupancy differences were associated with different histone post-translational modifications and combinatorial patterns of these marks were also associated with specific nucleosome organizations (Fig 2D–2F and S5 Table)[20, 41]. These findings, taken together with work that has shown variation in nucleosome repeat length is associated with different histone post-translational modifications, leads us to speculate that an important function of histone post-translational modifications could be to alter nucleosome organization[17, 25].

DNA methylation represents the final epigenetic modification that we analyzed. Our findings confirmed work done in plants that has shown enrichment for methylated cytosines within nucleosomes and display periodicities within the nucleosome core particle (Fig 3A and 3B and S5 Table and S12 Fig)[43]. Most interestingly, is that all four types of methylations are associated with distinct location preferences within the core particle. It will be interesting going forward to determine if these location preference differences have important functional significance, for example being used to fine tune nucleosome location during differentiation. Additionally, the CG methylation preferences around ± 40 bp most directly tie in methylation to increasing nucleosome occupancy since the three locations that have the highest potential nucleosome DNA binding capacity, dyad and ± 40bp, are the three highest enriched for CG methylations[44]. This is further corroborated by our findings of increasing nucleosome occupancy with an increasing number of methylations found within the core particle. These findings suggest that a role of DNA methylation is to alter nucleosome organization by increasing nucleosome occupancy, which can lead to deactivation. However, it must be stated that patterns of DNA methylations may not cause changes in nucleosome organization but could instead result from it.

We finally turned our attention toward determining if genome-wide nucleosome maps in combination with computational motif discovery tools alone were enough to determine which transcription factors are active in a cell type. It has been known for some time that DNase I hypersensitivity sites, which correspond to nucleosome free regions, can be used to find possible transcription factor binding sites. We sought to ask if nucleosome organization was sufficient to find these sites. By utilizing nucleosome architecture alone, our computational approach correctly predicted that the master regulators of pluripotency (OCT4, Nanog, Sox2, and KLF4) are active in H1 hESCs (Fig 4). Additionally, our approach also found other potential transcription factors that could be active in H1 stem cells, such as Tcf12, Mef2c, and HNF6 (S13 Fig). It will be interesting to see how much of a role, if any, these other factors play in maintaining pluripotency. Furthermore, this approach represents a proof of concept and further work can be done to fine-tune this approach. Since it has been shown that both DNase I and MNase maps introduce their own biases, it would be an interesting follow up application to integrate these two maps with motif analysis to discern if this approach could lead to a more robust platform for transcription factor discovery.

Methods

Cell culture

The UC Irvine Human Stem Cell Research Oversight Committee (UCI hSCRO) approved the use of human embryonic stem cells in this study. H1 and H9 human embryonic stem cell lines were purchased from WiCell Research Institute, Inc. These are some of the first ever human embryonic stem cell lines ever derived and are approved by the NIH Human Embryonic Stem Cell Registry (http://grants.nih.gov/stem_cells/registry/current.htm)[2]. The NIH Registration Numbers for H1 and H9 human embryonic stem cells are 0043 and 0062, respectively. Feeder free cultures of H1 and H9 human embryonic stem cells were grown and passaged in mTeSR 1 (STEMCELL Technologies Inc) as previously described and in accordance with ENCODE protocols to ease comparison to published ENCODE datasets[38]. In total, approximately 100 million H1 and H9 cells corresponding to passages 33–35 were used in subsequent experiments.

Generation of mono-nucleosomal DNA sequenced reads

H1 and H9 cells were subjected to MNase digestion by use of the EZ Nucleosomal DNA Kit (Zymo Research) in accordance with the manufacturer's protocol. The ideal digestion should

yield approximately 80% mono-nucleosomal DNA[14–17]. In order to extract both easily digested nucleosomes and less digestible ones, we titrated the time of digestion in multiple replicates to yield 70% to 90% mono-nucleosomal DNA, with the average being 80% from all replicates combined. We then prepared paired-end libraries from this total mono-nucleosomal DNA with use of the Illumina Paired-End DNA Sample Prep Kit according to the manufacturer's instructions with one exception. In order to reduce potential PCR amplification bias, we performed two separate PCR reaction steps and combined the product of the two reactions[42, 51]. The libraries were then sequenced using PE54 chemistry on the Illumina HiSeq2000 in replicate on two flow cells (R51 and R54). Two biological replicates for both H1 and H9 were performed, each consisting of six technical replicates for H1 and two technical replicates for H9.

Alignment and processing of nucleosome maps

Alignment. Paired-end nucleosomal sequencing data from H1 and H9 cells from R54 was aligned to the hg19 reference genome using the Bowtie2 algorithm on default settings[52]. Data from R51 was processed similarly with the exception that 25 bases from the 3' end were removed as these final cycles produced low Q-scores which caused excess reads to not align. Additionally, we processed raw sequencing paired-end nucleosome data on the same default settings using Bowtie2 for GM18507, GM18508, GM18516, GM18522, GM19193, GM19238, GM19239[23]. Raw sequencing data for the IV dataset was downloaded and aligned in colorspace by of default settings on Bowtie[17, 53].

Processing. All aligned data was processed using SAMtools to yield merged BAM files [54]. Finally, processed BAM files for K562 and GM12878 were downloaded from the ENCODE portal on the UCSC genome browser and merged[55].

Validation. BAM files from each H1 and H9 lane of sequencing data, as well as the merged datasets were compared using the deepTools bamCorrelate function[56]. The settings for bamCorrelate were as follows:—binSize 100,—corMethod Pearson,—outFileCorMatrix [H1 Table | H9 Table | H1-H9 Table],—plotFileFormat [H1 Heatmap | H9 Heatmap | H1-H9 Heatmap] (S2 and S3 Tables, S2 and S3 Figs).

Nucleosome occupancy score map generation and calling nucleosomes

BAM files were run through the DANPOS algorithm in which reads were clonally cut to remove potential PCR amplification bias, smoothed, and adjusted for nucleosome size to enhance signal to noise ratio, resulting in a nucleosome occupancy score (NOS) for each base in the human genome[37]. DANPOS settings were as follows for all paired-end datasets (H1, H9, GM18507, GM18508, GM18516, GM18522, GM19193, GM19238, GM19239):-d 150,-a 1,-k 1,-e 1,—paired 1. For single-read datasets (*in vitro*, K562, GM12878), the following DANPOS settings were used:-d 150,-a 1,-e 1,-k 1.-d 150 denoted setting the minimal distance between nucleosome dyads to 150 bp. The distance between dyads was set to 150 bp as the average fragment size from our H1 paired-end sequencing dataset was 151 bp (corresponding to 75 bp on either side of a dyad).-a 1 set the resolution of the NOS maps at a single bp and thus obviated any further downstream signal smoothing. The setting-e 1 allows for an edge-finding step to be taken, which estimates the edges of the predicted nucleosomes.-k 1 led to all data from intermediate steps being saved.—paired 1 indicated that the input BAM files were from paired-end sequencing data. This single base pair resolution NOS map was used to call best-fit nucleosomes with a corresponding average NOS so long as there was a minimum distance of 150 bp between two nucleosomes. Additionally, for all comparisons between different nucleosomal datasets, the NOS were normalized. We also generated NOS and called

nucleosomes for the H1 dataset corrected for MNase digestion bias with use of a genomic control and found no significant differences in sequence preference analyses (data not shown)[22, 23, 43, 57]. For all subsequent analyses we used our original NOS map and called nucleosomes.

General software used for analysis

Operations on genomic intervals were performed using BEDTools[58]. Fast Fourier transforms were done using MATLAB. Statistics were done in R. Heatmaps were generated with the Gitools software package[59]. Additionally, we made use of in-house Python 2.7, C++, and shell scripts that are available upon request.

Genomic annotations, DNA sequence preferences, clustering and total RNA-Signal generation

Mononucleotide and dinucleotide frequencies were computed by use of custom made Python and C++ programs. Gene coordinates were based on RefSeq coordinates that had at least 50% overlap with the consensus coding sequence (CCDS) gene coordinates[60, 61]. Additionally, we restricted the analysis to genes with unique transcription start sites, removing any duplicates. K-means clustering was performed as previously described[62, 63]. We initially chose a wide range of k values (data not shown) and used 10 as it yielded the clearest differences between clusters. Processed RNA sequence data was downloaded from ENCODE as bigWig files for the plus and minus strands[38]. The two signal files were normalized and added up to generate a single RNA-Signal file that we subsequently used to calculate transcriptionally silent genes and quantify total RNA-Expression of each gene from our initial list. Genome-wide Pearson's correlation coefficients (PCC) were performed by binning the NOS sets every 10 bp after removing coordinates corresponding to structural variants in H1 as defined by ENCODE[38].

Transcription factor binding, histone modifications, and chromatin states

Exon-exon junctions were downloaded from ENCODE[38]. We used ENCODE Affymetrix exon microarray data as an independent test to verify exon inclusion and exclusion in H1 transcripts. All enhancer and promoter coordinates were downloaded from ENCODE as were transcription factor binding sites and DNase-Seq signal. We called active enhancers and promoters as those that fell within one transcription factor binding site by ChIP-Seq data, had a high DNase-Seq signal and a low NOS. Encode histone-modification called peaks were used along with our called nucleosomes to assign called nucleosomes to one of ten corresponding modifications. Chromatin state start sites were downloaded from the UCSC genome-browser table.

DNA methylation analysis

Called and processed methylation data was downloaded and converted to hg19 using the liftOver utility[39, 40, 55]. Additionally, all sites called as methylated cytosines that were subsequently shown to be hydroxymethylcytosines were removed from the methylcytosine coordinates.

DNA motif analysis

Motif analysis was performed using HOMER with the following settings: Size 200, S 50, Len 6–14, Mis 3[45]. The input motif locations were determined by scanning the genome for the visual enhancer binding site motif (Fig 2C). Briefly, this is done by taking called nucleosomes from the DANPOS algorithm, and utilizing the NOS scores to locate the most well positioned nucleosomes with an intervening depleted region, that are also flanked by multiple additional well positioned nucleosomes, that also contain intervening depleted regions. The position of

the intervening depleted region of the most well positioned nucleosomes was then analyzed user HOMER.

Supporting Information

S1 Fig. Nucleosomal DNA post-MNase treatment. A, MNase treated DNA resolved in a 2% agarose gel stained with ethidium bromide and visualized with a UV light source. **B**, MNase treated DNA resolved in a 2% agarose gel stained with ethidium bromide and visualized with a visible blue light (DarkReader Transilluminator).
(TIF)

S2 Fig. Heatmap of H1 BAM comparison. Heatmap visualization with hierarchical clustering of the Pearson correlation coefficients performed for the H1 cell line as per S2 Table.
(TIF)

S3 Fig. Heatmap of H9 BAM comparison. Heatmap visualization with hierarchical clustering of the Pearson correlation coefficients performed for the H9 cell line as per S3 Table.
(TIF)

S4 Fig. Distribution of nucleosome fragment sizes. Fragment sizes were inferred through the use of the DANPOS algorithm.
(TIF)

S5 Fig. Mono- and Dinucleotide frequency analysis for H1. A, Mononucleotide frequencies in relation to the dyad. **B**, Dinucleotide frequencies for AA, TT, CC, and GG in relation to the dyad. **C**, Dinucleotide frequencies for all 16 dinucleotides plotted in four panels (top to bottom). **D**, Fast Fourier transforms (FFT) for all dinucleotides (minus AA and TT, see Fig 1A) plotted in four panels (top to bottom).
(TIF)

S6 Fig. Mono- and Dinucleotide frequency analysis for H9. A, Mononucleotide frequencies in relation to the dyad. **B**, Dinucleotide frequencies for AA, TT, CC, and GG in relation to the dyad. **C**, Dinucleotide frequencies for all 16 dinucleotides plotted in four panels (top to bottom). **D**, Fast Fourier transforms (FFT) for all dinucleotides plotted in five panels (top to bottom).
(TIF)

S7 Fig. NOS in association with gene features. A, Nucleosome occupancy scores (NOS) for all datasets (legend on top left) around transcription termination sites (TTS) from our 13,912 gene list used in all analyses. **B**, Codon start sites (CSS). **C**, Codon termination sites (CTS). **D**, Exon start sites (ESS). **E**, Exon termination sites (ETS).
(TIF)

S8 Fig. Heatmap of *k*-means clustered groups. Heatmaps of the H1 nucleosome occupancy scores (NOS) around transcription start sites (TSS) grouped by *k*-means clustering. The NOS for each location was divided by the maximum NOS. The median value is 0.201. Expression legend at top and bottom, with arrows denoting TSS.
(TIF)

S9 Fig. NOS around *k*-means clustered groups. Nucleosome occupancy scores (NOS) for all 12 datasets around 8 of 10 clusters, see Fig 1D for two others, grouped by *k*-means clustering of the H1 signal around transcription start sites (TSS), see S8 Fig for heatmaps of the groups.
(TIF)

S10 Fig. Site of maximum NOS for each cell line per cluster. Location of maximum NOS for all 12 cell lines per cluster around the TSS. Boxes are labelled as per the numbered clusters in Fig 1D and S9 Fig, as defined in S8 Fig, and indicate the region of the maximum NOS for the majority of cell lines for a given cluster.
(TIF)

S11 Fig. NOS around chromatin state start sites. On top is a table of chromatin state definitions for the 15 states used in Fig 2F. On bottom, nucleosome occupancy scores (NOS) of the H1 dataset around 15 chromatin state start sites grouped into panels based on similar functional candidate annotations.
(TIF)

S12 Fig. FFT of DNA methylation within the core particle. Fast Fourier transforms (FFT) of methylation frequencies within the nucleosome core particle with color coded legend on top.
(TIF)

S13 Fig. Transcription factors identified by motif analysis. HOMER identified transcription factor binding motifs within enhancer binding sites, as defined by surrounding nucleosome occupancies (Methods). Their associated gene, along with their q-value is also included.
(TIF)

S1 Table. Raw sequencing data metrics. Tables for both H1 and H9 (from top to bottom, respectively) with raw read count and alignment data from each biological replicate, as per Bowtie2.
(DOC)

S2 Table. Matrix of H1 bamCorrelate results. Table of Pearson correlation coefficients for next-generation sequencing data for the H1 cell line. This analysis compares all individual H1 replicates to one another, to each sequencing run, and to the pooled H1 dataset. Pooled datasets from the biological replicates are compared to one another and to the pooled H1 dataset.
(DOC)

S3 Table. Matrix of H9 bamCorrelate results. Table of Pearson correlation coefficients for next-generation sequencing data for the H9 cell line. This analysis compares all individual H9 replicates to one another, each sequencing run and to the pooled H9 dataset. Pooled datasets from the biological replicates are compared to one another and to the pooled H9 dataset.
(DOC)

S4 Table. *P*-values of post-translational modifications. Table of *p*-values generated by Mann-Whitney-Wilcoxon of the effect of histone post-translational modifications on nucleosome occupancy.
(DOC)

S5 Table. *P*-values for methylations. Top panel; Mann-Whitney-Wilcoxon *p*-values for all comparisons of the number of 5-methylcytosines found in a nucleosome and its effect on the average nucleosome occupancy. Bottom panel; Mann-Whitney-Wilcoxon *p*-values for all comparisons of the number of 5-hydroxymethylcytosines found in a nucleosome and its effect on the average nucleosome occupancy.
(DOC)

Acknowledgments

for help on Illumina sequencing technology (UCI Genomics High-Throughput Facility), and Bogi Andersen for reading the manuscript.

Author Contributions

References

1. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibro-blast cultures by defined factors. Cell. 2006; 126(4):663–76. doi: 10.1016/j.cell.2006.07.024 PMID: 16904174.

2. Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, et al. Embryonic stem cell lines derived from human blastocysts. Science. 1998; 282(5391):1145–7. PMID: 9804556.

3. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, et al. Genome-scale DNA methyl-ation maps of pluripotent and differentiated cells. Nature. 2008; 454(7205):766–70. doi: 10.1038/nature07107 PMID: 18600261; PubMed Central PMCID: PMC2896277.

4. Polo JM, Anderssen E, Walsh RM, Schwarz BA, Nefzger CM, Lim SM, et al. A molecular roadmap of reprogramming somatic cells into iPS cells. Cell. 2012; 151(7):1617–32. doi: 10.1016/j.cell.2012.11.039 PMID: 23260147; PubMed Central PMCID: PMC3608203.

5. Gifford CA, Ziller MJ, Gu H, Trapnell C, Donaghey J, Tsankov A, et al. Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. Cell. 2013; 153(5):1149–63. doi: 10.1016/j.cell.2013.04.037 PMID: 23664763; PubMed Central PMCID: PMC3709577.

6. Koche RP, Smith ZD, Adli M, Gu H, Ku M, Gnirke A, et al. Reprogramming factor expression initiates widespread targeted chromatin remodeling. Cell stem cell. 2011; 8(1):96–105. doi: 10.1016/j.stem.2010.12.001 PMID: 21211784; PubMed Central PMCID: PMC3220622.

7. Polo JM, Liu S, Figueroa ME, Kulalert W, Eminli S, Tan KY, et al. Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nature biotechnology. 2010; 28(8):848–55. doi: 10.1038/nbt.1667 PMID: 20644536; PubMed Central PMCID: PMC3148605.

8. Laurent LC, Ulitsky I, Slavin I, Tran H, Schork A, Morey R, et al. Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. Cell stem cell. 2011; 8(1):106–18. doi: 10.1016/j.stem.2010.12.003 PMID: 21211785; PubMed Central PMCID: PMC3043464.

9. Ma H, Morey R, O'Neil RC, He Y, Daughtry B, Schultz MD, et al. Abnormalities in human pluripotent cells due to reprogramming mechanisms. Nature. 2014; 511(7508):177–83. doi: 10.1038/nature13551 PMID: 25008523.

10. Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, et al. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell stem cell. 2010; 6(5):479–91. doi: 10.1016/j.stem.2010.03.018 PMID: 20452322; PubMed Central PMCID: PMC2867844.

11. Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, et al. Epigenomic analysis of multilineage dif-ferentiation of human embryonic stem cells. Cell. 2013; 153(5):1134–48. doi: 10.1016/j.cell.2013.04.022 PMID: 23664764; PubMed Central PMCID: PMC3786220.

12. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature. 1997; 389(6648):251–60. doi: 10.1038/38444 PMID: 9305837.

13. Segal E, Fondufe-Mittendorf Y, Chen L, Thastrom A, Field Y, Moore IK, et al. A genomic code for nucle-osome positioning. Nature. 2006; 442(7104):772–8. doi: 10.1038/nature04979 PMID: 16862119; PubMed Central PMCID: PMC2623244.

14. Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, et al. Dynamic regulation of nucleosome positioning in the human genome. Cell. 2008; 132(5):887–98. doi: 10.1016/j.cell.2008.02.022 PMID: 18329373.

15. Mavrich TN, Jiang C, Ioshikhes IP, Li X, Venters BJ, Zanton SJ, et al. Nucleosome organization in the Drosophila genome. Nature. 2008; 453(7193):358–62. doi: 10.1038/nature06929 PMID: 18408708; PubMed Central PMCID: PMC2735122.

16. Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, et al. The DNA-encoded nucle-osome organization of a eukaryotic genome. Nature. 2009; 458(7236):362–6. doi: 10.1038/nature07667 PMID: 19092803; PubMed Central PMCID: PMC2658732.

17. Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A. Determinants of nucleosome organization in primary human cells. Nature. 2011; 474(7352):516–20. doi: 10.1038/nature10002 PMID: 21602827; PubMed Central PMCID: PMC3212987.

18. Yen K, Vinayachandran V, Batta K, Koerber RT, Pugh BF. Genome-wide nucleosome specificity and directionality of chromatin remodelers. Cell. 2012; 149(7):1461–73. doi: 10.1016/j.cell.2012.04.036 PMID: 22726434; PubMed Central PMCID: PMC3397793.

19. Schuster SC, Miller W, Ratan A, Tomsho LP, Giardine B, Kasson LR, et al. Complete Khoisan and Bantu genomes from southern Africa. Nature. 2010; 463(7283):943–7. doi: 10.1038/nature08795 PMID: 20164927.

20. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011; 473(7345):43–9. doi: 10.1038/nature09906 PMID: 21441907; PubMed Central PMCID: PMC3088773.

21. Tolstorukov MY, Volfovsky N, Stephens RM, Park PJ. Impact of chromatin structure on sequence variability in the human genome. Nature structural & molecular biology. 2011; 18(4):510–5. doi: 10.1038/nsmb.2012 PMID: 21399641; PubMed Central PMCID: PMC3188321.

22. Brogaard K, Xi L, Wang JP, Widom J. A map of nucleosome positions in yeast at base-pair resolution. Nature. 2012; 486(7404):496–501. doi: 10.1038/nature11142 PMID: 22722846.

23. Gaffney DJ, McVicker G, Pai AA, Fondufe-Mittendorf YN, Lewellen N, Michelini K, et al. Controls of nucleosome positioning in the human genome. PLoS genetics. 2012; 8(11):e1003036. doi: 10.1371/journal.pgen.1003036 PMID: 23166509; PubMed Central PMCID: PMC3499251.

24. Li Z, Schug J, Tuteja G, White P, Kaestner KH. The nucleosome map of the mammalian liver. Nature structural & molecular biology. 2011; 18(6):742–6. doi: 10.1038/nsmb.2060 PMID: 21623366; PubMed Central PMCID: PMC3148658.

25. Teif VB, Vainshtein Y, Caudron-Herger M, Mallm JP, Marth C, Hofer T, et al. Genome-wide nucleosome positioning during embryonic stem cell development. Nature structural & molecular biology. 2012; 19(11):1185–92. doi: 10.1038/nsmb.2419 PMID: 23085715.

26. Zhang Z, Wippo CJ, Wal M, Ward E, Korber P, Pugh BF. A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. Science. 2011; 332(6032):977–80. doi: 10.1126/science.1200508 PMID: 21596991.

27. Gkikopoulos T, Schofield P, Singh V, Pinskaya M, Mellor J, Smolle M, et al. A role for Snf2-related nucleosome-spacing enzymes in genome-wide nucleosome organization. Science. 2011; 333(6050):1758–60. doi: 10.1126/science.1206097 PMID: 21940898; PubMed Central PMCID: PMC3428865.

28. Kundaje A, Kyriazopoulou-Panagiotopoulou S, Libbrecht M, Smith CL, Raha D, Winters EE, et al. Ubiquitous heterogeneity and asymmetry of the chromatin environment at regulatory elements. Genome research. 2012; 22(9):1735–47. doi: 10.1101/gr.136366.111 PMID: 22955985; PubMed Central PMCID: PMC3431490.

29. Peckham HE, Thurman RE, Fu Y, Stamatoyannopoulos JA, Noble WS, Struhl K, et al. Nucleosome positioning signals in genomic DNA. Genome research. 2007; 17(8):1170–7. doi: 10.1101/gr.6101007 PMID: 17620451; PubMed Central PMCID: PMC1933512.

30. Mavrich TN, Ioshikhes IP, Venters BJ, Jiang C, Tomsho LP, Qi J, et al. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome research. 2008; 18(7):1073–83. doi: 10.1101/gr.078261.108 PMID: 18550805; PubMed Central PMCID: PMC2493396.

31. Chen K, Wilson MA, Hirsch C, Watson A, Liang S, Lu Y, et al. Stabilization of the promoter nucleosomes in nucleosome-free regions by the yeast Cyc8-Tup1 corepressor. Genome research. 2013; 23(2):312–22. doi: 10.1101/gr.141952.112 PMID: 23124522; PubMed Central PMCID: PMC3561872.

32. Ioshikhes I, Hosid S, Pugh BF. Variety of genomic DNA patterns for nucleosome positioning. Genome research. 2011; 21(11):1863–71. doi: 10.1101/gr.116228.110 PMID: 21750105; PubMed Central PMCID: PMC3205571.

33. Kaplan N, Moore I, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, et al. Nucleosome sequence preferences influence in vivo nucleosome organization. Nature structural & molecular biology. 2010; 17(8):918–20. doi: 10.1038/Nsmb0810–918 PMID: WOS:000280670300002.

34. Zhang Y, Moqtaderi Z, Rattner BP, Euskirchen G, Snyder M, Kadonaga JT, et al. Evidence against a genomic code for nucleosome positioning Reply. Nature structural & molecular biology. 2010; 17(8):920–3. doi: 10.1038/Nsmb0810–920 PMID: WOS:000280670300003.

35. Zhang Y, Moqtaderi Z, Rattner BP, Euskirchen G, Snyder M, Kadonaga JT, et al. Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. Nature structural & molecular biology. 2009; 16(8):847–52. doi: 10.1038/nsmb.1636 PMID: 19620965; PubMed Central PMCID: PMC2823114.

36. Lantermann AB, Straub T, Stralfors A, Yuan GC, Ekwall K, Korber P. Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. Nature structural & molecular biology. 2010; 17(2):251–U15. doi: 10.1038/Nsmb.1741 PMID: WOS:000274228400021.

37. Chen K, Xi Y, Pan X, Li Z, Kaestner K, Tyler J, et al. DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. Genome research. 2013; 23(2):341–51. doi: 10.1101/gr.142067.112 PMID: 23193179; PubMed Central PMCID: PMC3561875.

38. Consortium EP, Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489(7414):57–74. doi: 10.1038/nature11247 PMID: 22955616; PubMed Central PMCID: PMC3439153.

39. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009; 462(7271):315–22. doi: 10.1038/nature08514 PMID: 19829295; PubMed Central PMCID: PMC2857523.

40. Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell. 2012; 149(6):1368–80. doi: 10.1016/j.cell.2012.04.027 PMID: 22608086; PubMed Central PMCID: PMC3589129.

41. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 2006; 125(2):315–26. doi: 10.1016/j.cell.2006.02.041 PMID: 16630819.

42. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, et al. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature. 2008; 452(7184):215–9. doi: 10.1038/nature06745 PMID: 18278030; PubMed Central PMCID: PMC2377394.

43. Chodavarapu RK, Feng S, Bernatavichute YV, Chen PY, Stroud H, Yu Y, et al. Relationship between nucleosome positioning and DNA methylation. Nature. 2010; 466(7304):388–92. doi: 10.1038/nature09147 PMID: 20512117; PubMed Central PMCID: PMC2964354.

44. Hall MA, Shundrovsky A, Bai L, Fulbright RM, Lis JT, Wang MD. High-resolution dynamic mapping of histone-DNA interactions in a nucleosome. Nature structural & molecular biology. 2009; 16(2):124–9. doi: 10.1038/nsmb.1526 PMID: 19136959; PubMed Central PMCID: PMC2635915.

45. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell. 2010; 38(4):576–89. doi: 10.1016/j.molcel.2010.05.004 PMID: 20513432; PubMed Central PMCID: PMC2898526.

46. Maherali N, Sridharan R, Xie W, Utikal J, Eminli S, Arnold K, et al. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell stem cell. 2007; 1(1):55–70. doi: 10.1016/j.stem.2007.05.014 PMID: 18371336.

47. Mikkelsen TS, Hanna J, Zhang X, Ku M, Wernig M, Schorderet P, et al. Dissecting direct reprogramming through integrative genomic analysis. Nature. 2008; 454(7200):49–55. doi: 10.1038/nature07056 PMID: 18509334; PubMed Central PMCID: PMC2754827.

48. Enroth S, Bornelov S, Wadelius C, Komorowski J. Combinations of histone modifications mark exon inclusion levels. PloS one. 2012; 7(1):e29911. doi: 10.1371/journal.pone.0029911 PMID: 22242188; PubMed Central PMCID: PMC3252363.

49. Schwartz S, Meshorer E, Ast G. Chromatin organization marks exon-intron structure. Nature structural & molecular biology. 2009; 16(9):990–5. doi: 10.1038/nsmb.1659 PMID: 19684600.

50. Tilgner H, Nikolaou C, Althammer S, Sammeth M, Beato M, Valcarcel J, et al. Nucleosome positioning as a determinant of exon recognition. Nature structural & molecular biology. 2009; 16(9):996–1001. doi: 10.1038/nsmb.1658 PMID: 19684599.

51. Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, et al. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome biology. 2011; 12(2):R18. doi: 10.1186/gb-2011-12-2-r18 PMID: 21338519; PubMed Central PMCID: PMC3188800.

52. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012; 9(4):357–9. doi: 10.1038/nmeth.1923 PMID: 22388286; PubMed Central PMCID: PMC3322381.

53. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology. 2009; 10(3):R25. doi: 10.1186/gb-2009-10-3-r25 PMID: 19261174; PubMed Central PMCID: PMC2690996.

54. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25(16):2078–9. doi: 10.1093/bioinformatics/btp352 PMID: 19505943; PubMed Central PMCID: PMC2723002.

55. Meyer LR, Zweig AS, Hinrichs AS, Karolchik D, Kuhn RM, Wong M, et al. The UCSC Genome Browser database: extensions and updates 2013. Nucleic acids research. 2013; 41(Database issue):D64–9. doi: 10.1093/nar/gks1048 PMID: 23155063; PubMed Central PMCID: PMC3531082.

56. Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic acids research. 2014; 42(Web Server issue):W187–91. doi: 10.1093/nar/gku365 PMID: 24799436; PubMed Central PMCID: PMC4086134.

57. Allan J, Fraser RM, Owen-Hughes T, Keszenman-Pereyra D. Micrococcal nuclease does not substantially bias nucleosome mapping. Journal of molecular biology. 2012; 417(3):152–64. doi: 10.1016/j.jmb.2012.01.043 PMID: 22310051; PubMed Central PMCID: PMC3314939.

58. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26(6):841–2. doi: 10.1093/bioinformatics/btq033 PMID: 20110278; PubMed Central PMCID: PMC2832824.

59. Perez-Llamas C, Lopez-Bigas N. Gitools: analysis and visualisation of genomic data using interactive heat-maps. PloS one. 2011; 6(5):e19541. doi: 10.1371/journal.pone.0019541 PMID: 21602921; PubMed Central PMCID: PMC3094337.

60. Pruitt KD, Harrow J, Harte RA, Wallin C, Diekhans M, Maglott DR, et al. The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. Genome research. 2009; 19(7):1316–23. doi: 10.1101/gr.080531.108 PMID: 19498102; PubMed Central PMCID: PMC2704439.

61. Pruitt KD, Tatusova T, Brown GR, Maglott DR. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic acids research. 2012; 40(Database issue):D130–5. doi: 10.1093/nar/gkr1079 PMID: 22121212; PubMed Central PMCID: PMC3245008.

62. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, et al. TM4 microarray software suite. Methods in enzymology. 2006; 411:134–93. doi: 10.1016/S0076–6879(06)11009–5 PMID: 16939790.

63. Soukas A, Cohen P, Socci ND, Friedman JM. Leptin-specific patterns of gene expression in white adipose tissue. Genes & development. 2000; 14(8):963–80. PMID: 10783168; PubMed Central PMCID: PMC316534.