DNA methylation marks inter-nucleosome linker regions throughout the human genome

Benjamin P. Berman[1,2,3,#], Yaping Liu[1,4], Theresa K. Kelly[1,2,*]

1) USC Epigenome Center

2) USC/Norris Comprehensive Cancer Center

3) Division of Bioinformatics, Department of Preventive Medicine

4) Graduate Program in Genetic, Molecular and Cellular Biology

University of Southern California, Los Angeles, CA

 # Corresponsing author

** Current address: Active Motif, Carlsbad, CA*

Contact info:

Benjamin P. Berman (bberman@usc.edu),

Yaping Liu (yapingli@usc.edu)

USC Epigenome Center

1450 Biggy St., #G511

Los Angeles, CA 90033

Ph: 323-442-7820

Terry Kelly (kelly@activemotif.com)

1914 Palomar Oaks Way, Suite 150

Carlsbad, CA 92008

Ph: 760-431-1263

PeerJ PrePrints | https://peerj.com/preprints/27v3/ | v3 received: 11 Jun 2013, published: 11 Jun 2013, doi: 10.7287/peerj.preprints.27v3

Background

Nucleosome organization and DNA methylation are two mechanisms that are important for proper control of mammalian transcription, as well as epigenetic dysregulation associated with cancer. Whole-genome DNA methylation sequencing studies have found that methylation levels in the human genome show periodicities of approximately 190 bp, suggesting a genome-wide relationship between the two marks. A recent report [1] attributed this to higher methylation levels of DNA within nucleosomes. Here, we analyzed a number of published datasets and found a more compelling alternative explanation, namely that methylation levels are highest in linker regions between nucleosomes.

Results

Reanalyzing the data from [1], we found that nucleosome-associated methylation could be strongly confounded by known sequence-related biases of the next-generation sequencing technologies. By accounting for these biases and using an unrelated nucleosome profiling technology, NOMe-seq, we found that genome-wide methylation was actually highest within linker regions occurring between nucleosomes in multi-nucleosome arrays. This effect was consistent among several methylation datasets generated independently using two unrelated methylation assays. Linker-associated methylation was most prominent within long Partially Methylated Domains (PMDs) and the positioned nucleosomes that flank CTCF binding sites. CTCF adjacent nucleosomes retained the correct positioning in regions completely devoid of CpG dinucleotides, suggesting that DNA methylation is not required for proper nucleosomes positioning.

Conclusions

The biological mechanisms responsible for DNA methylation patterns outside of gene promoters remain poorly understood. We identified a significant genome-wide relationship between nucleosome organization and DNA methylation, which can be used to more accurately analyze and understand the epigenetic changes that accompany cancer and other diseases.

Keywords: 5-methylcytosine, DNA methylation, nucleosome positioning, epigenetics, chromatin, CTCF

PeerJ PrePrints | https://peerj.com/preprints/27v3/ | v3 received: 11 Jun 2013, published: 11 Jun 2013, doi: 10.7287/peerj.preprints.27v3

Background

Packaging of DNA by nucleosomal proteins is an essential property of chromatin organization, and the precise positioning of individual nucleosomes at regulatory elements including promoters [2], enhancers [3], and insulators [4] is important for proper gene regulation [5]. Methylation of DNA at CpG dinucleotides also plays an important role in the regulation of transcription in mammals, and recent work has shown dynamic methylation changes occur at these same regulatory elements [6-9]. There is an intense interest in these two marks given that the genes controlling their deposition and removal are among the most commonly mutated in cancers [10, 11].

Recent advances in DNA sequencing have facilitated the production of maps covering the entire genome at single nucleotide resolution for both nucleosome positioning [2] and DNA methylation [12], yet the relationship between the two is poorly understood. In plants, methylation between cytosines in the CHG context was correlated at intervals of 175 base pairs, strongly suggesting an association with nucleosome positioning [13], but CHG methylation is not conserved in mammals. Comparing nucleosome positions genome-wide in plants and human embryonic stem cells showed a modest (roughly 2%) increase in DNA methylation over the nucleosome core, along with a 10bp periodicity that suggested methylation occurred specifically at positions where the major groove faced away from histone proteins [1]. More recently, *in vitro* nucleosome formation experiments showed that DNA methylation at the nucleosome core can promote the formation of a particular class of nucleosomes [14].

All of these earlier studies relied on MNase sequencing to define nucleosome positions *in vivo* and *in vitro*. Because MNase-seq and other "read enrichment" methods are known to introduce certain biases related to G/C content and other sequence composition [15-17], we developed a technique that does not depend on read enrichment to determine nucleosome positions, but rather uses a methyltransferase footprinting method [18]. NOMe-seq is based on bisulfite sequencing, and is therefore internally controlled for PCR and other steps that create skewed biases in read enrichment. We used NOMe-seq to investigate well-positioned arrays of nucleosomes surrounding CTCF binding sites, and discovered that DNA methylation was approximately two-fold higher in linker regions between nucleosomes than it was within the nucleosomes themselves [18]. This

PeerJ PrePrints | https://peerj.com/preprints/27v3/ | v3 received: 11 Jun 2013, published: 11 Jun 2013, doi: 10.7287/peerj.preprints.27v3

association with linker DNA was much stronger than the association reported previously for nucleosomal DNA [1], prompting us to re-analyze existing data in an attempt to reconcile these two results. It is worth noting that the two seemingly opposite associations are not mutually exclusive; methylation could be highest within linkers for some genomic elements, and highest in nucleosomes for others.

Results and Discussion

We first performed the same analysis of [1], aligning HSF1 embryonic stem cell DNA methylation levels to all MNase fragments from a CD4+ T-cell library [2]. This showed the same roughly 2% increase in methylation levels over the fragments, along with a clear 10-bp periodicity (Figure 1a). Reasoning that a deproteinated ("naked") DNA control would be completely devoid of *in vivo* nucleosome positioning information, we repeated the same analysis using a control library of naked HeLa DNA generated by the ENCODE project [19] (Figure 1a, pink lines). This data was generated by whole-genome sequencing of completely deproteinated genomic DNA that was fragmented by sonication. Methylation patterns aligned to these control fragments showed similar methylation patterns as the alignments to MNase based nucleosome fragments, suggesting a potential technical effect. We examined G/C content and found that fragments of both libraries were G/C rich, a factor known to introduce bias during the amplification involved in next-generation sequencing [17]. Why this G/C richness would cause higher methylation levels is not entirely understood, but it could be caused by a concomitant enrichment of CpG dinucleotides. While the mechanism is not understood, it is known that local CpG density is positively correlated with DNA methylation level ([20] and Additional File 1).

In an effort to identify nucleosome localization genome-wide without the potential influence of G/C content skew associated with individual sequencing fragments, we investigated the patterns of arrays of adjacent nucleosomes. It is clear from auto alignment of the MNase data that multi-nucleosome arrays are present throughout the genome (Figure 1b). We looked at methylation within an expanded region surrounding nucleosomes in whole-genome bisulfite sequencing (WGBS) data for cell types generated by different labs, including H1 [12] and HSF1 embryonic stem cells [1], IMR90 fibroblasts [12], normal and tumor colon tissue [21], and B-lymphocytes [22] (Figure 1c).

PeerJ PrePrints | https://peerj.com/preprints/27v3/ | v3 received: 11 Jun 2013, published: 11 Jun 2013, doi: 10.7287/peerj.preprints.27v3

Importantly, we included a dataset that was generated with a non-bisulfite approach, Methylation Sensitive Restriction Enzyme (MSRE) sequencing, to rule out any technical bisulfite effects. In all WGBS datasets, increased methylation was observed over MNase fragments. In both HSF1 [1] and IMR90 [12], this pattern was similar to the pattern for the naked DNA control (Figure 1c, right panel). When examining methylation levels outside the fragment itself, patterns in the MNase data diverged from the naked DNA control. All libraries except the most highly methylated hESC libraries showed increased methylation in inter-nucleosome linker regions (Figure 1c, left panel), supporting the relationship we had earlier observed in IMR90 nucleosomes adjacent to CTCF sites [18]. This relationship was strongest for the MSRE library, indicating a generality across cell types and methylation assays.

Next, we used the same analysis described above to investigate linker-specific IMR90 methylation in different genomic contexts. We were interested to see if methylated linkers were more prominent between nucleosomes positioned by CTCF binding sites as found previously [18], or within Partially Methylated Domains (PMDs) which have more variable methylation levels than the rest of the genome [12]. Indeed, linkers within PMDs and near CTCF sites were more strongly methylated than within non-PMDs (Figure 2a). CTCF regions showed the most dramatic linker-specific methylation, perhaps because they are the most consistently positioned class of nucleosomes in the genome. While the region immediately overlapping MNase fragments had strongly biased sequence composition, linker regions between nucleosomes had no sequence composition bias in any of the genomic contexts (Figure 2b). To validate genome-wide linker methylation, we identified consistent linker regions from IMR90 NOMe-seq nucleosome occupancy data [18] (Figure 2c). DNA within the linkers was consistently more methylated than the flanking nucleosomes, most prominently in CTCF regions and PMDs. Interestingly, in both MNase and NOMe-seq analysis, the inter-nucleosome spacing was shorter in CTCF regions (185bp) than PMDs or the rest of the genome (200bp). Genome-wide, we found that PMDs contained the bulk of all detectable nucleosomal periodicity (Figure 3).

To demonstrate that increased methylation in linker DNA was not cell type specific, we examined methylation around CTCF sites in several additional WGBS datasets as well as the non-bisulfite MSRE dataset described above. Indeed, all cell types showed linker-

PeerJ PrePrints | https://peerj.com/preprints/27v3/ | v3 received: 11 Jun 2013, published: 11 Jun 2013, doi: 10.7287/peerj.preprints.27v3

specific methylation (Figure 4a), and almost identical global patterns have been observed for dozens of other human tissues sequenced by WGBS in our lab (unpublished and data not shown). Interestingly, whereas CpGs within +/- 200bp of the CTCF binding site were completely unmethylated in most tissues, H1 and HSF1 embryonic stem cells (hESCs) showed increased methylation, possibly attributable to ESC-specific 5-hydroxymethylation at CTCF sites [23]. MSRE could not accurately represent the methylation levels within this +/- 200bp region due to known limitations of the method to measure very low methylation [22].

The large number of CTCF binding sites in the genome provided an opportunity to investigate the interplay between methylation and nucleosome positioning. There is evidence suggesting that methylation can influence nucleosome formation [14] and vice-versa [24]. It is impossible to determine with certainty without additional experiments, but we reasoned that if DNA methylation were required for nucleosome positioning, CpGs dinucleotides would be required around functional CTCF sites. To investigate this bioinformatically, we extracted CTCF-adjacent positions that contained zero CpGs in the reference human genome within a region of two full nucleosomes (+/-370bp). According to MNase occupancy and NOMe-seq chromatin accessibility levels, the nucleosomes at these "zero CpG" regions were positioned just as well as other CTCF-adjacent nucleosomes, strongly suggesting that linker DNA methylation is not necessary for nucleosome positioning (Figure 4b-c). Nevertheless, the "zero CpG" regions comprise only about 1-3% of CTCF-adjacent nucleosomes, so we can not completely rule out some role for DNA methylation in establishing or reinforcing nucleosome positioning.

Conclusion

We have provided strong evidence for a pervasive methylation pattern occurring at linker regions between arrays of positioned nucleosomes in the human genome. This observation has implications for methylome analysis, suggesting that methylation levels may be used to deduce nucleosome positioning in some cases. Nucleosomes adjacent to CTCF binding sites may account for a significant fraction of these nucleosomal arrays, since it is estimated that approximately one million nucleosomes may be positioned adjacent to CTCF sites (around 55,000 CTCF sites in any given cell type [7], with about 20

PeerJ PrePrints | https://peerj.com/preprints/27v3/ | v3 received: 11 Jun 2013, published: 11 Jun 2013, doi: 10.7287/peerj.preprints.27v3

nucleosomes positioned per site [4]). We additionally showed that methylation levels within linker regions are unlikely to play a causal role in the positioning of CTCF-adjacent nucleosomes. This is parsimonious with the observation that strongly positioned nucleosomes are stacked against a barrier introduced by ATP-dependent nucleosome remodeling [25]. These observations also fit well with data showing that CTCF binding precedes local demethylation of CTCF binding sites [6].

Inhibition of DNA methylation has been demonstrated for certain histone modifications, including H3K4me1,2,3 [24] and H2A.Z [26]. Because CTCF-adjacent nucleosomes are marked by both of these modifications, it is attractive to hypothesize that inhibition by these modifications does not extend into the linker regions, leaving them open to DNA methyltransferase activity. We did observe significant nucleosomal periodicity in regions outside of known CTCF sites (data not shown), and we found that the bulk of this periodicity was within PMD regions (Figure 3), which are depleted for active histone marks such as H3K4me1,2,3 and H2A.Z. The higher level of nucleosomal periodicity detected within PMDs may be a consequence the high methylation state maintained outside of PMDs [27]. Further analysis is necessary to identify precise histone modification of nucleosomes and methylation status in the same reference cell type.

Finally, based on our observations of methylation patterns within MNase and naked DNA sequencing fragments, we also suggest that appropriate controls are necessary for MNase-seq to rule out small biases introduced by next-generation sequencing. G/C content and MNase-specific cleavage biases are known to be difficult confounders of MNase-seq [28, 29], and we have proposed NOMe-seq [18] as a complementary strategy that can be used to validate any results that might be affected by sequence-specific biases.

Materials and Methods

CpG methylation datasets: Percent methylation was taken from WGBS supplemental data files from Lister et al. [12] (IMR90, H1) and Berman et al. [21] (tumor colon and normal, [GEO:GSE32399]). For B-lymphocyte MSRE dataset, supplemental data files from Ball et al. [22] contained the number of tag counts for each possible HspII site. Using the procedure described in the Ball et al. “methods” section, we transformed these counts to

PeerJ PrePrints | https://peerj.com/preprints/27v3/ | v3 received: 11 Jun 2013, published: 11 Jun 2013, doi: 10.7287/peerj.preprints.27v3

percent methylation using the following equation: **m** = 1 – (0.1124 * **c**), where **m** is the estimated percent methylation, and **c** is the raw tag counts.

IMR90 NOMe seq data: NOMe-seq data was taken from Kelly et al. [18] [GEO:GSE40770]. A beta-binomial Hidden Markov Model (HMM) [30] was used to identify linker regions (manuscript in preparation).

IMR90 MNase-seq (figure 4 only): IMR90 cells were cultured according to ATCC's guidelines. Mononucleosomes were generated by digesting $1x10^6$ cells with 0.5, 1 and 5 Units of micrococcal nuclease (MNase; Worthington Biochemicals) for 15 minutes at 37 °C. The three MNase preparations were combined, and mononucleosome fragments of ~150 bp were gel extracted and libraries were preparared from 30ng DNA using Illumina single-end sequencing adapters as described in [31]. Sequencing was performed on an Illumina Genome Analyzer IIx using standard Illumina reagents, producing 153,469,077 high quality 36bp sequence reads. Reads were aligned using MAQ with a minimum mapping quality of 30, resulting in 111,705,730 uniquely alignable reads. All sequences and alignments are available at [GEO:GSE21823].

Nucleosome occupancy score (Figure 4 only): For genomic coordinate **c** and an estimated mononucleosome size **s**, the nucleosome occupancy score for a particular position was determined by summing the number of MNase tags on the forward genomic strand in the range **c**-(**s**/2) and the number of tags on the reverse strand in the range **c**+(**s**/2). We estimated **s** to be 165 after examining a range of values (50bp-250bp) within 1kb of all CTCF binding sites. After alignment to the genomic element of interest, the raw nucleosome occupancy score was normalized for local tag density by dividing by the total number of reads within 200bp. Plots were smoothed by taking a moving average of normalized occupancy scores within a 20bp window.

CTCF datasets: CTCF binding sites were taken from [21]. For "CTCF regions with 0 CpGs", we used only those genomic positions that contained no CpGs in the reference

PeerJ PrePrints | https://peerj.com/preprints/27v3/ | v3 received: 11 Jun 2013, published: 11 Jun 2013, doi: 10.7287/peerj.preprints.27v3

human genome within a span of two nucleosomes on either side (+/-370bp). This comprised about 1% of the full CTCF set.

PeerJ PrePrints | https://peerj.com/preprints/27v3/ | v3 received: 11 Jun 2013, published: 11 Jun 2013, doi: 10.7287/peerj.preprints.27v3

Competing Interests

Dr. Kelly is one of several inventors listed on a patent related to the NOMe-seq technology (US20120028817). More than a year after her contribution to the work described here, Dr. Kelly became an employee of Active Motif (Carlsbad, CA), which markets a NOMe-seq kit.

Author Contributions

BPB and YL performed data analysis. TKK performed MNase-seq experiments and helped interpret results. BPB conceived the study, supervised the work, and wrote the manuscript.

Acknowledgements

We wish to thank Peter Jones for his suggestions and for fostering a supportive and collaborative environment, and Huy Dinh, Clayton Collings, and Fides Lay for helpful comments on the manuscript. We are grateful to Charlie Nicolet, Selene Tyndale, and Helen Truong of the Norris Molecular Genomics Core at the USC Epigenome Center for generating the sequencing data. Computation was performed at the USC High Performance Computing and Communications Center (http://www.usc.edu/hpcc/). This project described was supported in part by T32CA00932027 from the National Cancer Institute to TKK, and the generous support of the Kenneth T. and Eileen L. Norris Foundation to YL and BPB. It was also supported in part by award number P30CA014089 from the National Cancer Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

PeerJ PrePrints | https://peerj.com/preprints/27v3/ | v3 received: 11 Jun 2013, published: 11 Jun 2013, doi: 10.7287/peerj.preprints.27v3

References

1. Chodavarapu RK, Feng S, Bernatavichute YV, Chen P-Y, Stroud H, Yu Y, Hetzel Ja, Kuo F, Kim J, Cokus SJ, et al: **Relationship between nucleosome positioning and DNA methylation.** *Nature* 2010, **466:**388-392.
2. Schones DE, Cui K, Cuddapah S, Roh T-Y, Barski A, Wang Z, Wei G, Zhao K: **Dynamic regulation of nucleosome positioning in the human genome.** *Cell* 2008, **132:**887-898.
3. He HH, Meyer Ca, Shin H, Bailey ST, Wei G, Wang Q, Zhang Y, Xu K, Ni M, Lupien M, et al: **Nucleosome dynamics define transcriptional enhancers.** *Nature genetics* 2010, **42:**343-347.
4. Fu Y, Sinha M, Peterson CL, Weng Z: **The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome.** *PLoS genetics* 2008, **4:**e1000138.
5. Iyer VR: **Nucleosome positioning: bringing order to the eukaryotic genome.** *Trends in cell biology* 2012, **22:**250-256.
6. Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, et al: **DNA-binding factors shape the mouse methylome at distal regulatory regions.** *Nature* 2011, **480:**490-495.
7. Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, Lee K, Canfield T, Weaver M, Sandstrom R, et al: **Widespread plasticity in CTCF occupancy linked to DNA methylation.** *Genome research* 2012, **22:**1680-1688.
8. Gifford CA, Ziller MJ, Gu H, Trapnell C, Donaghey J, Tsankov A, Shalek AK, Kelley DR, Shishkin AA, Issner R, et al: **Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells.** *Cell* 2013, **153:**1149-1163.
9. Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, Whitaker JW, Tian S, Hawkins RD, Leung D, et al: **Epigenomic Analysis of Multilineage Differentiation of Human Embryonic Stem Cells.** *Cell* 2013, **153:**1134-1148.
10. You JS, Jones Pa: **Cancer genetics and epigenetics: two sides of the same coin?** *Cancer cell* 2012, **22:**9-20.
11. Dawson Ma, Kouzarides T: **Cancer epigenetics: from mechanism to therapy.** *Cell* 2012, **150:**12-27.
12. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo Q-m, et al: **Human DNA methylomes at base resolution show widespread epigenomic differences.** *Nature* 2009, **462:**315-322.
13. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE: **Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning.** *Nature* 2008, **452:**215-219.
14. Collings CK, Waddell PJ, Anderson JN: **Effects of DNA methylation on nucleosome stability.** *Nucleic acids research* 2013**:**1-14.
15. Dohm JC, Lottaz C, Borodina T, Himmelbauer H: **Substantial biases in ultra-short read data sets from high-throughput DNA sequencing.** *Nucleic acids research* 2008, **36:**e105.

PeerJ PrePrints | https://peerj.com/preprints/27v3/ | v3 received: 11 Jun 2013, published: 11 Jun 2013, doi: 10.7287/peerj.preprints.27v3

16. Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S, Frazer Ka: **Evaluation of next generation sequencing platforms for population targeted sequencing studies.** *Genome biology* 2009, **10:**R32.
17. Benjamini Y, Speed TP: **Summarizing and correcting the GC content bias in high-throughput sequencing.** *Nucleic acids research* 2012, **40:**e72.
18. Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, Jones Pa: **Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules.** *Genome research* 2012, **22:**2497-2506.
19. Auerbach RK, Euskirchen G, Rozowsky J, Lamarre-Vincent N, Moqtaderi Z, Lefrançois P, Struhl K, Gerstein M, Snyder M: **Mapping accessible chromatin regions using Sono-Seq.** *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106:**14926-14931.
20. Edwards JR, O'Donnell AH, Rollins Ra, Peckham HE, Lee C, Milekic MH, Chanrion B, Fu Y, Su T, Hibshoosh H, et al: **Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns.** *Genome research* 2010, **20:**972-980.
21. Berman BP, Weisenberger DJ, Aman JF, Hinoue T, Ramjan Z, Liu Y, Noushmehr H, Lange CPE, van Dijk CM, Tollenaar RaEM, et al: **Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains.** *Nature genetics* 2012, **44:**40-46.
22. Ball MP, Li JB, Gao Y, Lee J-H, LeProust EM, Park I-H, Xie B, Daley GQ, Church GM: **Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells.** *Nature biotechnology* 2009, **27:**361-368.
23. Yu M, Hon GC, Szulwach KE, Song C-X, Jin P, Ren B, He C: **Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine.** *Nature protocols* 2012, **7:**2159-2170.
24. Ooi SKT, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin S-P, Allis CD, et al: **DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA.** *Nature* 2007, **448:**714-717.
25. Zhang Z, Wippo CJ, Wal M, Ward E, Korber P, Pugh BF: **A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome.** *Science (New York, NY)* 2011, **332:**977-980.
26. Zilberman D, Coleman-Derr D, Ballinger T, Henikoff S: **Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks.** *Nature* 2008, **456:**125-129.
27. Raddatz G, Gao Q, Bender S, Jaenisch R, Lyko F: **Dnmt3a protects active chromosome domains against cancer-associated hypomethylation.** *PLoS genetics* 2012, **8:**e1003146.
28. Zhang Y, Moqtaderi Z, Rattner BP, Euskirchen G, Snyder M, Kadonaga JT, Liu XS, Struhl K: **Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo.** *Nature structural & molecular biology* 2009, **16:**847-852.

PeerJ PrePrints | https://peerj.com/preprints/27v3/ | v3 received: 11 Jun 2013, published: 11 Jun 2013, doi: 10.7287/peerj.preprints.27v3

29. Chung H-R, Dunkel I, Heise F, Linke C, Krobitsch S, Ehrenhofer-Murray AE, Sperling SR, Vingron M: **The effect of micrococcal nuclease digestion on nucleosome positioning data.** *PloS one* 2010, **5:**e15754.
30. Molaro A, Hodges E, Fang F, Song Q, McCombie WR, Hannon GJ, Smith AD: **Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates.** *Cell* 2011, **146:**1029-1041.
31. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, et al: **A bivalent chromatin structure marks key developmental genes in embryonic stem cells.** *Cell* 2006, **125:**315-326.

PeerJ PrePrints | https://peerj.com/preprints/27v3/ | v3 received: 11 Jun 2013, published: 11 Jun 2013, doi: 10.7287/peerj.preprints.27v3

Figure Legends

Figure 1: **Methylation levels relative to MNase-seq fragments.** (A) Including an additional control to the analysis performed by Chodavarapu et al. [1] shows that HSF1 methylation levels are increased over the MNase fragments from the CD4+ T-cell dataset used in Chodavarapu et al. (red line), but are also increased over fragments from a whole-genome sequencing library generated by sonication of deproteinated ("naked") genomic DNA (pink line). The right panel shows elevated G/C content levels over these same fragments. (B) Alignment of MNase cut sites relative to MNase fragments reveals ordered arrays of nucleosomes, suggesting pervasive nucleosomal arrays genome-wide. (C) Various WGBS methylation levels are aligned to MNase (left) and Naked DNA (right) fragments, along with a methylation library generated with non-bisulfite MSRE sequencing (see text). Elevated methylation levels are observed covering both MNase and Naked DNA fragments, but linker regions are elevated only relative to MNase library.

Figure 2: Increased methylation in linker regions within different genomic contexts. (A) IMR90 methylation patterns around MNase fragments were plotted as in Figure 1, but stratified by genomic context. IMR90 Partially Methylated Domains (PMDs) are from [12], while non-PMD contains the remainder of the genome. See the methods section for a description of CTCF binding sites. The left column shows methylation on a consistent scale, while the middle column zooms into a scale relevant for each context. (B) Local CpG density aligned to the same MNase fragments. (C) Linkers identified from IMR90 NOMe-seq [18] are shown aligned to IMR90 chromatin accessibility (GCH, green line) and methylation (HCG, black line). H can include any A, C, or T nucleotide.

Figure 3: Linker-specific methylation is higher within PMDs. (B) Concordance between nearby CpGs. This was defined as the fraction of reads that were methylated at a given CpG, plotted as a function of the genomic distance from a

PeerJ PrePrints | https://peerj.com/preprints/27v3/ | v3 received: 11 Jun 2013, published: 11 Jun 2013, doi: 10.7287/peerj.preprints.27v3

reference methylated CpG (mCpG). If the target CpG had multiple reference mCpGs within 2kb interval, it was counted separately for each.

Figure 4: **DNA methylation occurs primarily at linker regions in nucleosomal arrays flanking CTCF binding sites.** (A) Methylation levels around motifs bound by CTCF in HeLa cells (see methods). Association between methylation and nucleosome positioning is verified in several WGBS datasets and one non-bisulfite (MSRE) dataset. (B) Nucleosome occupancy is shown around CTCF sites for IMR90 cells. The black line includes all CTCF-adjacent regions from Figure 4a. The red line includes only positions that have zero CpGs within +/-370 base pairs (a region the size of four full nucleosomes). (C) Same analysis, but using NOMe-seq chromatin accessibility from IMR90 cells [18].

Additional File 1 (PDF): Genome-wide correlation between local CpG density and DNA methylation. (A) Data from IMR90 cells [12] was extracted from all non-overlapping 100bp bins on chr17, and ranked by CpG density. Groups of 100 bins were averaged to show CpG density, CpG methylation, and tag density for H3K4me3 ChIP-seq. At CpGs without K4me3 mark, increasing local CpG density is correlated with DNA methylation level. (B) The reason for this is unknown, but this is an agreement with an earlier study of human breast and brain tissues [20].

PeerJ PrePrints | https://peerj.com/preprints/27v3/ | v3 received: 11 Jun 2013, published: 11 Jun 2013, doi: 10.7287/peerj.preprints.27v3

FIG 1

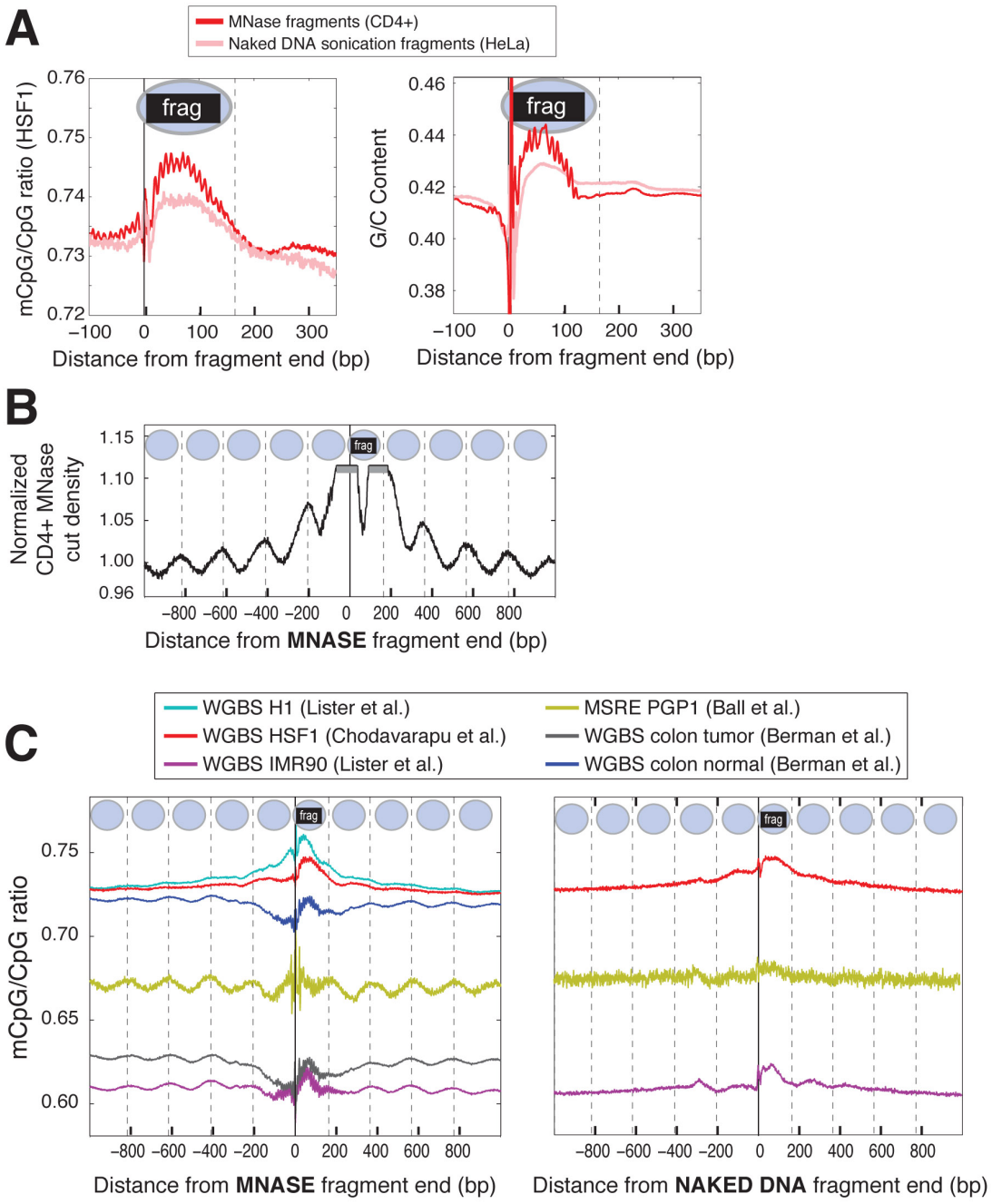


Figure 1: Methylation levels relative to MNase-seq fragments. (A) Including an additional control to the analysis performed by Chodavarapu et al. (Chodavarapu et al. 2010) shows that HSF1 methylation levels are increased over the MNase fragments from the CD4+ T-cell dataset used in Chodavarapu et al. (red line), but are also increased over fragments from a whole-genome sequencing library generated by sonication of deproteinated ("naked") genomic DNA (pink line). The right panel shows elevated G/C content levels over these same fragments. (B) Alignment of MNase cut sites relative to MNase fragments reveals ordered arrays of nucleosomes, suggesting pervasive nucleosomal arrays genome-wide. (C) Various WGBS methylation levels are aligned to MNase (left) and Naked DNA (right) fragments, along with a methylation library generated with non-bisulfite MSRE sequencing (see text). Elevated methylation levels are observed covering both MNase and Naked DNA fragments, but linker regions are elevated only relative to MNase library.

FIG 2

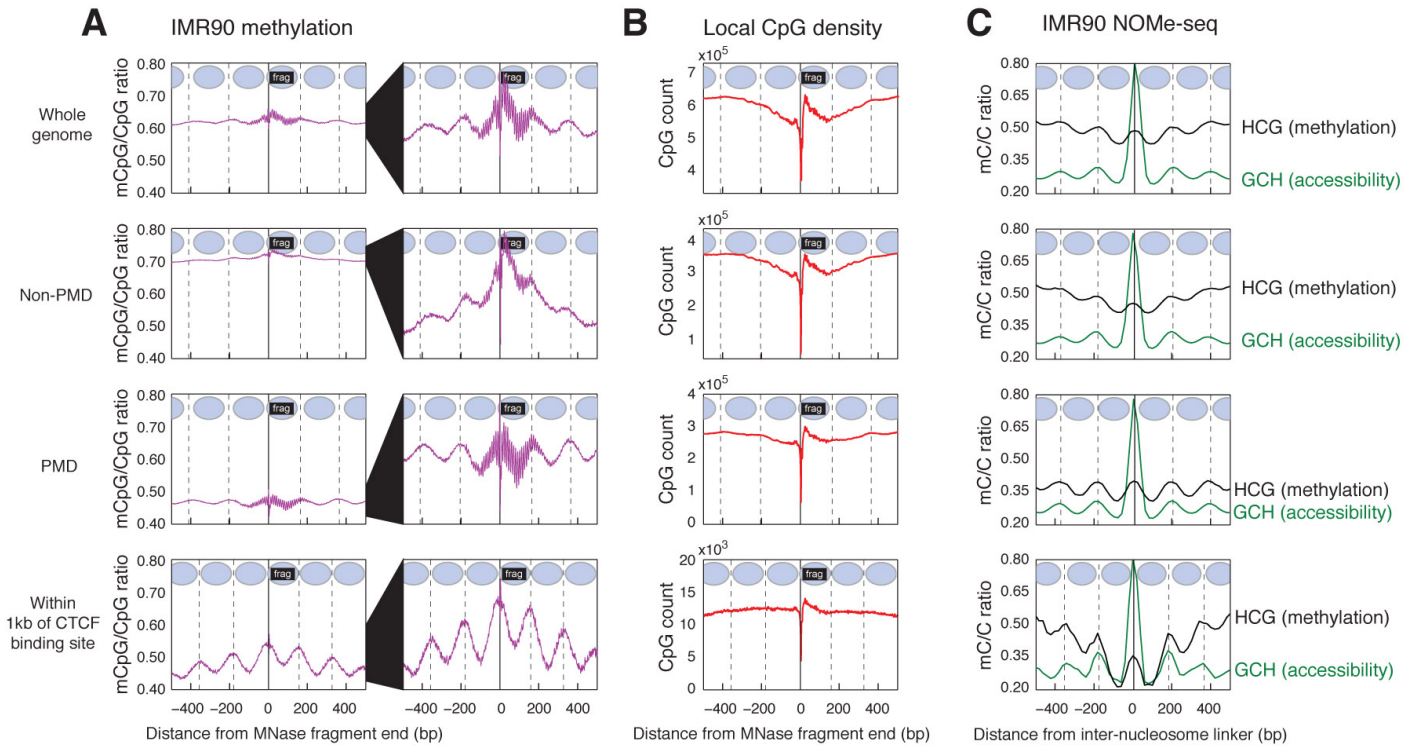


Figure 2: Increased methylation in linker regions within different genomic contexts. (A) IMR90 methylation patterns around MNase fragments were plotted as in Figure 1, but stratified by genomic context. IMR90 Partially Methylated Domains (PMDs) are from (Lister et al. 2009), while non-PMD contains the remainder of the genome. See the methods section for a description of CTCF binding sites. The left column shows methylation on a consistent scale, while the middle column zooms into a scale relevant for each context. (B) Local CpG density aligned to the same MNase fragments. (C) Linkers identified from IMR90 NOMe-seq (Kelly et al. 2012) are shown aligned to IMR90 chromatin accessibility (GCH, green line) and methylation (HCG, black line). H can include any A, C, or T nucleotide.

FIG 3

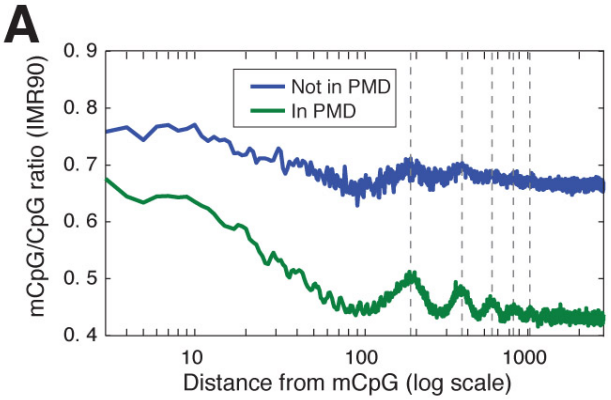


Figure 3: Linker-specific methylation is higher within PMDs. (B) Concordance between nearby CpGs. This was defined as the fraction of reads that were methylated at a given CpG, plotted as a function of the genomic distance from a reference methylated CpG (mCpG). If the target CpG had multiple reference mCpGs within 2kb interval, it was counted separately for each.

FIG 4

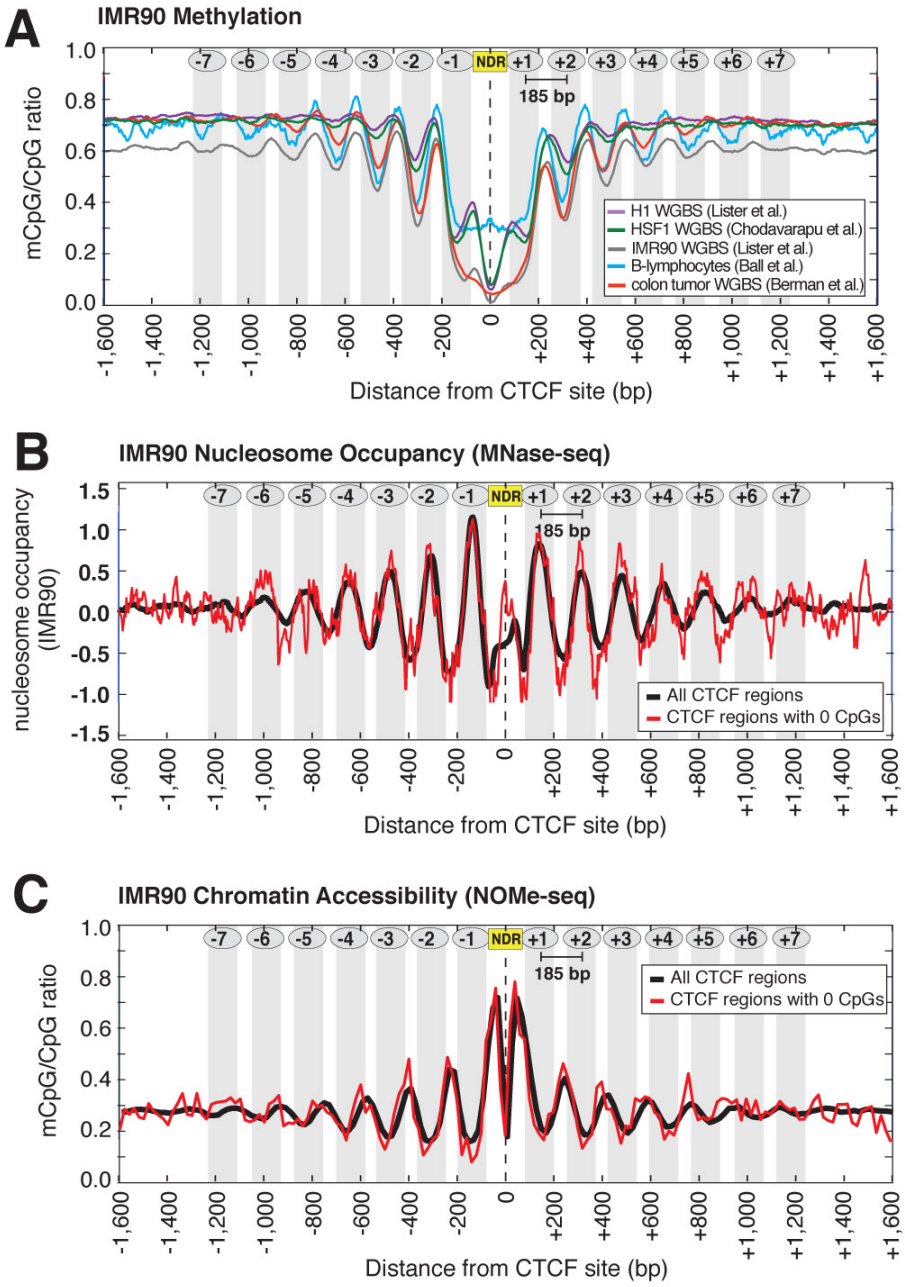
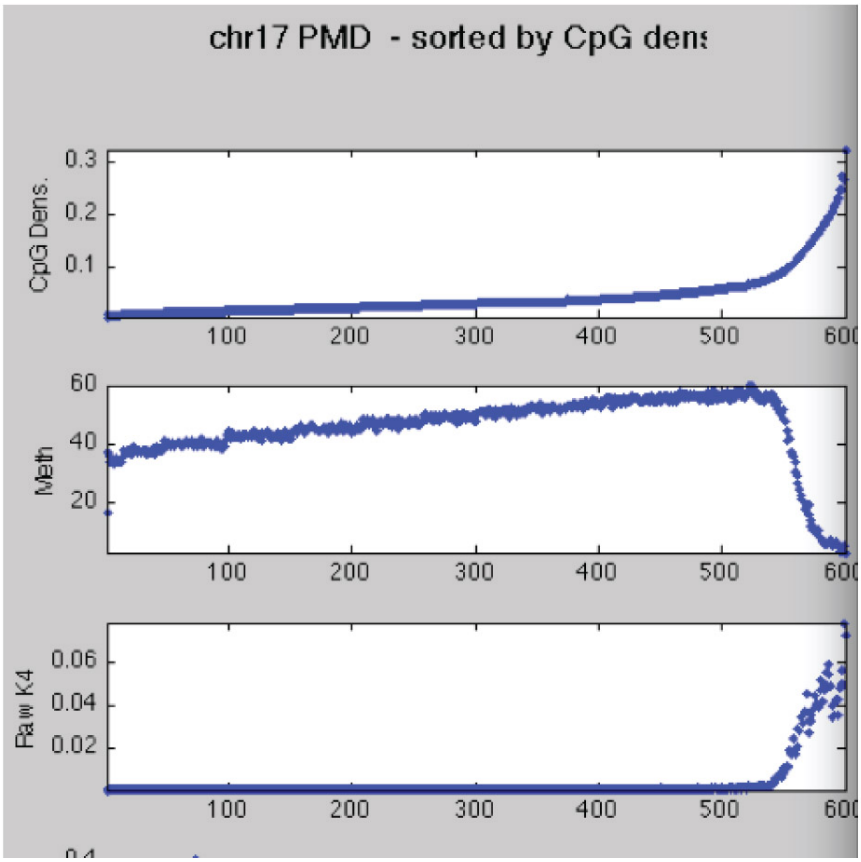


Figure 4: DNA methylation occurs primarily at linker regions in nucleosomal arrays flanking CTCF binding sites. (A) Methylation levels around motifs bound by CTCF in HeLa cells (see methods). Association between methylation and nucleosome positioning is verified in several WGBS datasets and one non-bisulfite (MSRE) dataset. (B) Nucleosome occupancy is shown around CTCF sites for IMR90 cells. The black line includes all CTCF-adjacent regions from Figure 4a. The red line includes only positions that have zero CpGs within +/-370 base pairs (a region the size of four full nucleosomes). (C) Same analysis, but using NOMe-seq chromatin accessibility from IMR90 cells (Kelly et al. 2012).

ADDITIONAL FILE 1



Supplemental Figure S1: Genome-wide correlation between local CpG density and DNA methylation. (A) Data from IMR90 cells (Lister et al. 2009) was extracted from all non-overlapping 100bp bins on chr17, and ranked by CpG density. Groups of 100 bins were averaged to show CpG density, CpG methylation, and tag density for H3K4me3 ChIP-seq. At CpGs without K4me3 mark, increasing local CpG density is correlated with DNA methylation level. (B) The reason for this is unknown, but this is an agreement with an earlier study of human breast and brain tissues (Edwards et al. 2010).