



Reconfiguration of nucleosome depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer

Phillippa Claire Taberlay, Aaron Luke Statham, Theresa Kathleen Kelly, et al.

Genome Res. published online June 10, 2014

Access the most recent version at doi:[10.1101/gr.163485.113](https://doi.org/10.1101/gr.163485.113)

Supplemental Material <http://genome.cshlp.org/content/suppl/2014/06/23/gr.163485.113.DC1.html>

P<P Published online June 10, 2014 in advance of the print journal.

Accepted Manuscript Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.

Open Access Freely available online through the *Genome Research* Open Access option.

Creative Commons License This manuscript is Open Access. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 3.0 Unported), as described at <http://creativecommons.org/licenses/by/3.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To subscribe to *Genome Research* go to:
<http://genome.cshlp.org/subscriptions>

Reconfiguration of Nucleosome Depleted Regions at Distal Regulatory Elements Accompanies DNA Methylation of Enhancers and Insulators in Cancer

Phillippa C. Taberlay^{1,2,3#}, Aaron L. Statham^{2,4#}, Theresa K. Kelly^{1,4}, Susan J. Clark^{2,3,5*} and Peter A. Jones^{1,5*}

¹. Departments of Biochemistry and Urology, Norris Comprehensive Cancer Center, University of Southern California, 1441 Eastlake Avenue, Los Angeles, CA 90033, USA

². Epigenetics Research, Cancer Program, Garvan Institute of Medical Research, 384 Victoria Street, Darlinghurst, NSW 2010 AUSTRALIA

³. St. Vincent's Clinical School, Faculty of Medicine, University of New South Wales, Darlinghurst, New South Wales 2010, AUSTRALIA

⁴. Active Motif, Inc., 1914 Palomar Oaks Way, Suite 150, Carlsbad, CA 92008, USA

⁵. Correspondence should be addressed to Peter A. Jones (p.jones@med.usc.edu) or Susan J. Clark (s.clark@garvan.org.au)

^{#,*} These authors contributed equally

KEYWORDS

Epigenome, DNA methylation, nucleosome deplete regions, enhancers, insulators, distal regulatory elements, cancer.

ABSTRACT

It is well established that cancer-associated epigenetic repression occurs concomitant with CpG island hypermethylation and loss of nucleosomes at promoters, but the role of nucleosome occupancy and epigenetic reprogramming at distal regulatory elements in cancer is still poorly understood. Here, we evaluate the scope of global epigenetic alterations at enhancers and insulator elements in prostate and breast cancer cells using simultaneous genome-wide mapping of DNA methylation and nucleosome occupancy (NOME-seq). We find that the genomic location of nucleosome depleted regions (NDRs) is mostly cell type specific and preferentially found at enhancers in normal cells. In cancer cells however, we observe a global reconfiguration of NDRs at distal regulatory elements coupled with a substantial reorganization of the cancer methylome. Aberrant acquisition of nucleosomes at enhancer-associated NDRs is associated with hypermethylation and epigenetic silencing marks and conversely, loss of nucleosomes with demethylation and epigenetic activation. Remarkably, we show that nucleosomes remain strongly organized and phased at many facultative distal regulatory elements, even in the absence of a NDR as an anchor. Finally we find that key transcription factor (TF) binding sites also show extensive peripheral nucleosome phasing, suggesting the potential for TFs to organize NDRs genome-wide and contribute to deregulation of cancer epigenomes. Together our findings suggests that “decommissioning” of NDRs and TFs at distal regulatory elements in cancer cells is accompanied by DNA hypermethylation susceptibility of enhancers and insulator elements which in turn may contribute to an altered genome-wide architecture and epigenetic deregulation in malignancy.

INTRODUCTION

Epigenetic mechanisms are necessary for normal cellular functions and are mitotically heritable through many cellular divisions. Disruption of faithful maintenance of epigenetic processes, including DNA methylation, histone composition, post-translational histone modifications and nucleosome occupancy (Coolen et al. 2010; Taberlay and Jones 2011; Bert et al. 2013), can lead to global changes to the cancer epigenome coupled with associated deregulation of cancer gene expression signatures. DNA methylation currently remains the most characterized epigenetic mechanism in normal and cancer biology. Seminal studies have demonstrated that while CpG island associated promoters are commonly hypermethylated and silenced (Jones and Baylin 2007; Baylin and Jones 2011), the bulk of the genome is abnormally hypomethylated in cancer cells. Genome-wide hypomethylation occurs commonly at intergenic regions and likely contributes to genomic instability during transformation (Almeida et al. 1993; Ji et al. 1997), highlighting the critical importance of the epigenome in determining chromatin structure.

The nucleosome is the organizational unit of chromatin, and can therefore be considered one of the underlying drivers of the epigenetic state and ultimately, transcriptional output. As such, it is increasingly important to consider how nucleosomes are organized in the context of the entire epigenome. Pioneering studies that couple nuclease-treated DNA with genome-wide sequencing have already provided a wealth of information regarding global nucleosome occupancy (Yuan et al. 2005; Ioshikhes et al. 2006; Lee et al. 2007; Schones et al. 2008). Notably, that nucleosomes are absent from the -1 position near transcriptional start sites of transcribed genes (Yuan et al. 2005; Ioshikhes et al. 2006; Lee et al. 2007; Lin et al.

2007; Whitehouse et al. 2007; Field et al. 2008; Schones et al. 2008; Shivaswamy et al. 2008; Valouev et al. 2008; Field et al. 2009), but are surrounded by highly-organised nucleosomes enriched in active histone modifications (Barski et al. 2007). We recently demonstrated that the strength of a promoter nucleosome depleted region (NDR) is correlated with the associated level of gene expression (Kelly et al. 2012). Yet, the physical presence of the nucleosome is also important; they anchor DNA methyltransferases (Jeong et al. 2009; Sharma et al. 2011) that are necessary for DNA methylation (You et al. 2011) and/or carry repressive histone modifications that contribute to formation of inaccessible chromatin structures (Taberlay et al. 2011). Indeed, the key role of the nucleosome is gaining recognition in both normal and malignant promoter regulation (Kelly et al. 2010; Wolff et al. 2010; Andreu-Vieyra et al. 2011; Taberlay et al. 2011; You et al. 2011). To date however, research has primarily focused on DNA methylation and loss of function of CpG island containing promoters, which represent only a small proportion of the potential regulatory nodes in cancer.

Distal regulatory elements such as enhancers and insulators determine the transcriptional profile of a cell in addition to promoters. Our recent work has revealed that inactive, but permissive, enhancers exhibit a remarkably similar epigenetic signature to that of active enhancers, and are commonly paired with promoters that are repressed by the Polycomb repressive complex in order to facilitate cellular reprogramming by master regulatory factors (Taberlay et al. 2011). In contrast, we found that cancer cells are rendered resistant to cellular reprogramming if the cognate enhancer is occupied by nucleosomes and methylated. These data suggested that

distal regulatory elements are also subject to epigenetic modification and could be a key feature of the cancer epigenome.

Here, we have integrated epigenome-wide maps of DNA methylomes and nucleosome occupancy in prostate and breast normal and cancer cells to gain a more comprehensive understanding of the relationship between NDRs and cancer related epigenetic changes that occur at distal regulatory regions. We find that reconfiguration of NDRs at distal regulatory elements in cancer cells is associated with epigenetic deregulation and a change in DNA methylation of enhancers and insulator elements.

RESULTS

Nucleosome depleted regions (NDRs) are largely unique to individual cell types.

We recently developed a method for high-resolution simultaneous genome-wide mapping of DNA methylation and nucleosome occupancy within individual DNA strands using nucleosome occupancy and methylome sequencing (NOMe-seq) (Kelly et al. 2012). NOMe-seq exploits the ability of *M.CviPI* (Xu et al. 1998) to methylate GpC sites within regions of DNA that are not occupied by nucleosomes or other tight-binding proteins (Jessen et al. 2004; Hoose and Kladde 2006; Jessen et al. 2006; Kilgore et al. 2007; Kelly et al. 2010; Andreu-Vieyra et al. 2011; Taberlay et al. 2011; You et al. 2011; You et al. 2013) (Figure 1). Thus, NOMe-seq produces a high-resolution digital readout of endogenous DNA methylation (from CpG sites) and nucleosome occupancy or depletion (from GpC sites) within individual molecules, which removes the need for a population-based analysis to infer a relationship between DNA methylation and nucleosome occupancy. Here, we performed the first comparative analyses of NOMe-seq data from multiple cell types, including normal human mammary epithelial cells (HMEC) and breast cancer cell line (MCF7), as well as normal prostate epithelial cells (PrEC) and prostate cancer cell line (PC3), to define the altered patterns of DNA methylation, histone modifications and nucleosome occupancy that occur at distal regulatory elements in two different cell line models of cancer.

First, we developed a modified version of our custom bisulfite sequencing analysis pipeline (Figure 1; see also Methods)(Statham et al. 2012) to call nucleosome-depleted regions (NDRs). These were defined as having a significant enrichment of GpC methylated sites above the background distribution across a minimum of 140bp,

independent of other genomic landmarks (e.g. transcriptional start sites), and called at increasing stringencies (range; low: $-\log_{10}$ p-value=5, high: $-\log_{10}$ p-value=20). We found that few NDRs were shared between the breast cell lines (HMEC and MCF7), with overlap being ~10% regardless of the cutoff used (Figure S1A). The same trend was observed when we compared PrEC and PC3 prostate cells (Figure S1B). Based on this analysis, we selected a mid-range $-\log_{10}$ p-value=15 for all subsequent experiments to ensure high-confidence in NDR calls.

To further investigate the extent of unique and common NDRs, we extended our comparison to include all four cell types. Interestingly, we found that 65.84% out of a total of 88,578 NDRs detected were unique to only one cell type (Figure S1C). We tested whether this was solely due to the stringency chosen and found that <4% of NDRs are shared between all four cell types, even at the highest stringency (Figure S1D). These data show a plateau at $-\log_{10}$ p-value=15 and that $\leq 20\%$ NDRs are shared between any two cell types, regardless of the pairs. We also intersected all NDRs to show the relationship between cell types (Figure S1E). Taken together our data demonstrate that nucleosome occupancy at a given genomic point is largely unique to individual cell types independent of normal or cancer status.

We next asked whether the NDR calls detected by NOME-seq overlapped accessible regions identified by DNase-seq and FAIRE-seq, which are commonly used techniques to detect the presence of accessible chromatin. A comparison of these assays revealed that NOME-seq, DNase-seq and FAIRE-seq each detect only a subset of accessible regions (Figure S2A). Notably, these disparities may be attributed to each technique having a different preference to GpC density (Figure S2B-D). DNase-

seq has a preference for GC-rich regions (Figure S2B), while FAIRE-seq interrogates GC-deplete regions (Figure S2C). In contrast, NOMe-seq spans both GC-deplete and GC-rich regions (Figure S2D).

NDRs are enriched at enhancers in normal epithelial cells, and change to predominantly overlap CTCF in cancer cells. To determine if NDRs were differentially located at functional genomic loci between normal and cancer cells, we needed to consider NDRs in the context of epigenome maps extending beyond DNA methylation and nucleosome occupancy using NOMe-seq. We therefore performed ChIP-seq to generate signatures of key histone modifications (H3K4me1, H3K27ac, H3K4me3, H3K27me3) and regulatory factors (CTCF, RNA Pol II) and then applied the multivariate Hidden Markov Model, ChromHMM (Ernst and Kellis 2010; Ernst et al. 2011; Ernst and Kellis 2012), to annotate the epigenomes of each normal and cancer cell type into nine distinct chromatin states (heterochromatin, repressed, transcribed, enhancers, enhancers + CTCF, CTCF, promoters + CTCF, promoters and poised_promoters (Figure S3A)). Each chromatin state was enriched with the expected complement of histone modifications (Figure S3A). Additionally, we confirmed that RNA-seq profiles surrounding each of these ChromHMM states were as expected (eg. HMEC and MCF7, Figure S3B).

We then intersected our NOMe-seq datasets with each of the ChromHMM states and asked where NDRs were most dominant in the genome (Figure 2). In HMEC, the majority of NDRs were found to overlap enhancers (50.12%) (Figure 2A). In contrast, only 20.34% of NDRs overlap an enhancer in MCF7 cancer cells. Instead, most NDRs cover CTCF sites (32.66%) in MCF7 cells (Figure 2A) compared to 9.77% that

overlap CTCF sites in HMEC cells (Figure 2A). However, the proportion of NDRs overlapping most other regulatory regions does not change between HMEC and MCF7 (Figure 2A). Importantly, the same trend is observed in the prostate cells, where NDR locations localize more commonly to enhancers in PrEC compared to CTCF occupied regions in PC3 (Figure 2B). We also performed the same genomic analysis, but using DNase I hypersensitive sites as the measure of accessibility (Figure 2C-D). We again observed fewer DNase I hypersensitive sites overlapping enhancers (Figure 2C-D) and a greater number of DNase I sites overlapping CTCF sites in cancer cells (Figure 2C-D), supporting the NOME-seq results (Figure 2A-B). Our data suggests that while NDRs are mostly unique to each cell type (Figure S1), the genomic locations of NDRs are commonly shared between normal cells which are predominantly at enhancers, whereas in cancer cells NDRs are located primarily at CTCF insulator elements.

Epigenetic silencing of regulatory elements extends to include both enhancers and insulators in cancer cells. Our previous work revealed that the cognate enhancer of *MYOD1* is devoid of H3K4me1 and occluded by nucleosomes in cancer cells (Taberlay et al. 2011). This prompted us to ask whether abnormal changes in nucleosome occupancy, histone marks and DNA methylation at distal regulatory elements occurs on a global scale and could be considered a general characteristic of cancer cells. To address this question, we selected the promoters, CTCF sites and enhancers that were defined in normal cells (ChromHMM; HMEC and PrEC) and then characterized the epigenetic changes that occur at these *precise* genomic regions in cancer cells (MCF7 and PC3). For example, enhancers defined in normal cells were separated into four categories based on their chromatin state in cancer cells: (1)

“retains both” (NDR maintained and enhancer marks maintained in cancer cells), (2) “loses enhancer” (NDR maintained, enhancer marks lost in cancer cells), (3) “loses NDR” (NDR lost, enhancer marks maintained in cancer cells) and (4) “loses both” (NDR lost, enhancer marks lost in cancer cells) (Figure 3A-B). The same categorization was also performed for promoters and CTCF sites. For each of these categories, we then assessed changes in DNA methylation to establish the spectrum of epigenetic alterations that can occur in cancer. At promoters, repressive epigenetic changes involved the concomitant loss of the H3K4me3 modification, loss of the NDR (that is, the acquisition of a nucleosome) and increased DNA methylation (Figure 3A, left panel). For example, a promoter overlapping *NFYC* exhibits a NDR and a complement of active histone marks in HMEC cells, but displays features of heterochromatin in MCF7 cells (Figure 3C). A similar pattern was observed at CTCF sites whereby epigenetic changes involved the concomitant loss of CTCF binding, acquisition of a nucleosome and increased DNA methylation (Figure 3A, middle panel). Figure 3C (middle panel) also shows an example of an intergenic insulator located on chromosome 12 where there is a loss of CTCF binding, a gain of a nucleosome and gain of DNA methylation in MCF7 cells relative to HMEC cells. These data suggest that epigenetic silencing requires loss of the NDR in concert with removal of H3K4me3 or CTCF from promoters and insulators in cancer cells.

In contrast, we found that enhancer epigenetic silencing can be distinct from both promoters and CTCF sites (Figure 3A, right panel). At enhancers, increased DNA methylation was observed in the presence of H3K4me1, but was associated with the gain of a nucleosome (“loses NDR”; Figure 3A, right panel). Therefore, increased DNA methylation at enhancers does not require removal of the H3K4me1 mark. The

same analysis in prostate cells demonstrated that these observations were not limited to breast cells (Figure 3B, right panel). Indeed, as we described in breast cancer cells, promoters and CTCF sites shared common features of epigenetic change, while enhancers displayed different behavior in prostate cancer cells (Figure 3B). Together, our data infers that enhancer epigenetic silencing, as exemplified by increased DNA methylation, is unique from promoters and insulators since the process does not necessarily require the removal of the defining mark (H3K4me1); suggesting that loss of the NDR, alone, can predict increased DNA methylation events at enhancers in cancer cells.

We sought to validate the genome-wide findings and used the NOME assay (Kelly et al. 2010; Andreu-Vieyra et al. 2011; Taberlay et al. 2011; You et al. 2011; You et al. 2013) to assess nucleosome occupancy changes at a distal regulatory element that were classified as an enhancer in normal cells but marked as heterochromatin in cancer cells (Figure 3D, upper panel). We interrogated nucleosome occupancy within individual enhancer modules in the two normal and two cancer cell types. As observed in the genome-wide data, a clear NDR is evident across the enhancer in both HMEC and PrEC cells (Figure 3D, lower panel). In cancer cells, all individual enhancer modules were occupied by nucleosomes (MCF7 and PC3) confirming that this region exhibits a closed chromatin structure, consistent with the “heterochromatin” ChromHMM state detected in the cancer cells. Together, these data support our finding that fewer NDRs overlap enhancers in cancer cells and is also consistent with our observation that a greater number of CTCF sites overlap a NDR in cancer cells (Figure 2).

Epigenetic activation can encompass distal gene regulatory elements. We then sought to determine whether a subset of promoters, CTCF sites and enhancers could conversely acquire an abnormally active epigenetic signature. We identified a subset of regulatory elements that either gained H3K4me3 (at promoters), and or CTCF binding (at insulators) or H3K4me1 (at enhancers). Indeed, we found that promoters, CTCF sites and enhancers could all exhibit characteristics of epigenetic activation in MCF7 breast cancer cells (Figure 4A) and in PC3 prostate cancer cells (Figure 4B). Interestingly, we noticed that fewer regulatory elements aberrantly acquire an active epigenetic signature; for example, 2142 enhancers gain an active signature (reduced DNA methylation, nucleosome loss (NDR formed) and increased H3K4me1; Figure 4A) while 6356 enhancers become silenced (increased DNA methylation, nucleosome gain (NDR lost) and removal of H3K4me1; Figure 3A) in MCF7 compared to HMEC. This trend is consistent with previous findings (Akhtar-Zaidi et al. 2012) and suggests that it is more difficult to activate a silenced regulatory element than it is to silence an already active or poised regulatory element.

NOMe-seq reveals that nucleosomes are phased throughout the genome, even at weak or absent Nucleosome Depleted Regions (NDRs). We next asked whether an underlying feature of the genome could determine the precise location of an NDR since our data points to nucleosome occupancy as a cell-type specific feature. We took advantage of the single-molecule readout of NOMe-seq to determine whether global nucleosome occupancy at a given genomic region is random, or organized. We examined nucleosome phasing ± 1 kb from (1) all NDRs in each cell type and (2) all other regions of the genome that could potentially exhibit a NDR (*facultative*; *fNDR*) as defined in at least one of the four NOMe-seq datasets, but not in the specific cell

type examined. At NDRs we observed phasing of at least four nucleosomes flanking the NDRs in HMEC and MCF7 cells. Notably, the nucleosome phasing matched an apparent methylation phasing with the peaks of DNA methylation occurring between the phased nucleosomes and low levels of DNA methylation at the NDR (Figure 5A-B). Hypomethylation at NDRs and similar exquisite phasing of flanking nucleosomes and methylation patterns was also observed in the PrEC and PC3 NOME-seq datasets (Figures S4A-B). This is consistent with the finding that unmethylated sequences are refractory to stable nucleosome formation (Collings et al. 2013).

Next, we mapped nucleosome phasing ± 1 kb from the center of *f*NDRs (i.e. nucleosome occupied in the cell type examined but depleted in one of the other cell types). In HMEC, the precise phasing of the nucleosomes at *f*NDRs was striking and unexpectedly consistent with the nucleosome organization observed at NDRs (compare broken teal line to solid teal line, Figure 5A). The phasing at *f*NDRs was again pronounced in MCF7 (Figure 5B), PrEC (Figure S4A) and PC3 (Figure S4B) confirming that this pattern was not unique to HMEC cells. To validate this unexpected finding, we next performed the same analysis using a $-\log_{10}$ p-value = 3 to more definitively separate NDRs and *f*NDRs. At this threshold, weak NDRs are removed but phasing is still observed in the “absent” bin (Figure S4G-J). Irrespective of threshold, nucleosomes flanking the *f*NDRs were also accompanied by a pattern of phased DNA methylation (Figures 5A-B and S4A-B).

Nucleosomes are not phased when DNase I hypersensitive sites are absent, or CTCF is not bound. DNase I hypersensitivity is a measure of accessible DNA that infers nucleosome depletion (Hogan et al. 2006; Giresi et al. 2007; Kim et al. 2007),

or lack of transcription factor binding (Song et al. 2011). Using a similar approach as for NDRs (Figures 5A-B), we analyzed nucleosome occupancy ± 1 kb from (1) all DNase I hypersensitivity sites in each cell type and (2) all other regions of the genome that have the potential to be DNase I hypersensitive (*facultative*; *f*DNase I) in at least one of the four cell types. We observed that nucleosomes were depleted from DNase I hypersensitive sites (Figures 5C-D and S4C-D) and DNA methylation was low (Figures 5C-D and S4C-D), as expected. Nucleosome phasing is evident surrounding DNase I sites in all four datasets but diminished at *f*DNase I sites (Figures 5C-D and S4C-D). Similarly, DNA methylation phasing is observed at DNase I and absent from *f*DNase I sites (Figures 5C-D and S4C-D). This is in contrast to DNA methylation phasing observed at both NDRs and *f*NDRs.

Nucleosomes are organized surrounding CTCF sites (Kelly et al. 2012; Wang et al. 2012b). Therefore, using a similar approach as for NDRs (Figure 5A-B) and DNase I (Figure 5C-D), we examined nucleosome phasing ± 1 kb from all CTCF-occupied regions in each cell type as established by ChIP-seq. We found that CTCF binding sites are unmethylated and nucleosome depleted and interestingly, display a small region of inaccessibility at the peak center, corresponding to a footprint of the bound CTCF complex (Figures 5E-F and S4E-F). As expected, nucleosomes and DNA methylation were strongly phased surrounding CTCF binding sites. However, a further ~5% dip in DNA methylation was observed at the peak of CTCF binding.

Whether nucleosome distribution is perturbed in the absence of CTCF binding has not been addressed. Therefore, we also examined all other regions of the genome that have the potential to be occupied by CTCF (*facultative*; *f*CTCF) in at least one of the

four cell types. At f CTCF sites, where CTCF is not bound, nucleosome phasing is not evident (Figures 5E-F and S4E-F). Consistent with a lack of organized nucleosomes (Figures 5E-F and S4E-F), f CTCF sites do not exhibit DNA methylation phasing (Figures 5E-F and S4E-F). Global DNA hypomethylation was evident in both MCF7 and PC3 cancer cells surrounding CTCF and f CTCF sites, as well as the NDRs and DNase I sites.

Transcriptional regulator binding sites are nucleosome depleted, hypomethylated and are associated with precise nucleosome phasing. In addition to CTCF, nucleosome phasing has also been detected at other transcriptional regulator binding sites (Wang et al. 2012b) raising the possibility that nucleosome phasing surrounding regulatory factors may be widespread. To determine the extent of nucleosome organization, we mapped nucleosome occupancy and DNA methylation adjacent to the peak center of twenty transcriptional regulators from publicly available MCF7 datasets (Supplementary Table 1). We examined nucleosome phasing ± 1 kb from each transcriptional regulator binding site and observed marked nucleosome depletion in all datasets (Figures 6A-C and S5A-B). DNA methylation was low (typically ~5%-10%) at the center of transcriptional regulator binding, despite many of these factors not containing CpG sites in their consensus sequence. Of the twenty transcriptional regulators we examined, only three were associated with strong nucleosome phasing – CTCF, REST and RAD21 (Figure 6A-C), where we observed the phasing of at least five nucleosomes flanking the NDR. This was accompanied by DNA methylation phasing, whose peak corresponded with the trough in the nucleosome array (consistent with linker regions; Figure 6A-C). Interestingly, small patches of inaccessibility corresponding to footprints created by CTCF, REST or

RAD21 binding and the associated dip in DNA methylation were only observed in the presence of strongly phased nucleosomes (Compare Figures 6A-F and S5A-B). In contrast ELF1, FOSL2, GABP, JUND, HDAC2 and MAX corresponded with weak nucleosome phasing and almost indiscernible phasing patterns in DNA methylation (Figure S5A), while nucleosome phasing was not detected at the remainder of the transcriptional regulator binding sites (Figure S5B). Taken together, these data suggest that only certain regulatory factors play key roles in establishing nucleosome and DNA methylation phasing patterns.

DISCUSSION

The critical events that underlie epigenetic aberrations at promoters have been largely deciphered (reviewed by (Jones and Baylin 2007; Baylin and Jones 2011; Taberlay and Jones 2011)); yet, epigenetic alterations encompassing nucleosome occupancy, histone marks and DNA methylation at distal regulatory regions has not been similarly detailed. We utilized comprehensive epigenome-wide maps of DNA methylomes and nucleosome occupancy (NOMe-seq) in cell line models of prostate and breast cancer to gain an understanding of nucleosome distribution in each cell type and define a spectrum of epigenetic changes that occur at distal regulatory regions. Here, we show that NDRs and key TFs play an important role in phasing nucleosomes and establishing flanking DNA methylation patterns. Moreover, we find that in cancer cells, “decommissioning” of NDRs and TFs at distal regulatory elements is associated with DNA hypermethylation of enhancers and insulator elements and gain of repressive epigenetic marks (Figure 7). In addition we show that loss of nucleosomes at distal regulatory elements is associated with demethylation and gain of active epigenetic marks (Figure 7). Our study emphasizes the importance of understanding epigenetic control in the context of the entire epigenome to determine how genomic circuits typified by distal regulator elements may be disrupted during malignancy.

Foremost, and irrespective of normal or cancer status, NOMe-seq methodology allowed us to observe that the phasing of nucleosomes occurs directly with phasing of DNA methylation, which is not possible with other assays that determine accessibility such as DNase-seq or FAIRE-seq. We observed that phasing occurs in all cell types although it remains unclear what is dictating the precise organization that is evident at

NDRs and even at absent NDRs (*f*NDRs). The first possibility is a sequence-based propensity for DNA to be associated with nucleosome occupancy (Bernstein et al. 2004; Segal et al. 2006; Kaplan et al. 2009; Kelly and Jones 2011; Hughes et al. 2012; Struhl and Segal 2013) at these regulatory elements, which supports a model whereby nucleosome organization may be more conserved than currently anticipated. Our data also points to a second scenario whereby select regulatory factors are also critical for organizing nucleosomes and subsequently, DNA methylation patterns. All of the transcription factor binding sites that we examined displayed a NDR that is conspicuously unmethylated. Yet, while the majority of transcription factors showed little or no evidence of nucleosome or peripheral DNA methylation phasing, nucleosome and DNA methylation phasing is highly pronounced around CTCF, REST and RAD21 binding sites. These sites display a patch of inaccessibility at the binding midpoint, consistent with a transcription factor footprint observed in previous reports using DNase-seq (Neph et al. 2012; Wang et al. 2012b). That DNA methylation is highly phased in concert with organized nucleosomes surrounding these transcription factor binding sites is consistent with the observation that linker regions are more accessible to DNA methyltransferases (DMTases) (Felle et al. 2011), as we have observed (Kelly et al. 2012) and shown to be due to “seeding” methylation that occurs preferentially between nucleosomes at linker regions and then spreads to nucleosomes (Hinshelwood et al. 2009). Interestingly, our data shows that the pattern of DNA methylation phasing is highly consistent with the strength of the nucleosome phasing, further highlighting the strong interaction between DMTases and nucleosomes. However, it should be noted that many transcription factor recognition sequences do not contain CpG sites and therefore, DNA methylation may not be directly influencing their binding profiles, as has been reported for MYC

(Prendergast et al. 1991; Prendergast and Ziff 1991). It will be important to consider in future studies whether the strength, stability or physical bulk of the transcription factor binding complexes and or the genomic sequence context can explain why only some regulatory factors are associated with highly organized nucleosomes flanking NDRs.

Nucleosome depleted enhancers are critical for the activity of master regulators that determine cellular phenotype (Taberlay and Jones 2011). Here we show that very few NDR locations are shared between cell-types; consistent with observations that the majority of enhancers are cell-type specific (Heintzman and Ren 2007; Visel et al. 2007; Heintzman et al. 2009). Categorization of the epigenome into ChromHMM states showed that NDRs are predisposed to overlap enhancers in normal cells; however, we found that NDRs preferentially overlapped CTCF sites in cancer cells. The question remains as to the underlying biological impact of this substantial re-organization of NDRs in cancer cells. It has been proposed that architectural proteins such as CTCF may be associated with widespread changes to the cancer methylome in addition to alterations in the physical packaging of chromatin (Berman et al. 2012), and perhaps phasing. Our results support this hypothesis, as we can only detect phasing when CTCF occupies its cognate binding site, whereas both nucleosome phasing and DNA methylation patterns are disrupted when CTCF is not bound. This suggests that CTCF and most likely RAD21 and other factors such as REST are driving the arrangement of nucleosomes and perhaps dictating DNA methylation patterns, offering a potential explanation for the substantial re-organization of NDRs to primarily overlap CTCF in the cancer cells.

This may be particularly important in the context of cancer where global hypomethylation occurs concomitant with aberrant, but targeted epigenetic silencing of discrete genomic regions such as promoters (Taberlay and Jones 2011). Indeed, it has long been accepted that epigenetic switching is a hallmark of promoters in cancer cells (*e.g.* (Gal-Yam et al. 2008)). We now show that epigenetic silencing of distal regulatory elements is achieved through increased DNA methylation, altered nucleosome occupancy and loss of characteristic signatures (such as H3K4me1 or CTCF that largely define enhancers or insulators, respectively). While current technologies limit our ability to dissect a specific order of events, we have previously demonstrated that nucleosomes first occupy enhancers, prior to DNA methylation events (You et al. 2011). Since enhancers are typically CpG-poor (Andreu-Vieyra et al. 2011; Stadler et al. 2011; Jones 2012), our data reveals that aberrant epigenetic silencing or activation of these distal regulatory elements is not directly governed by CpG density. Our analysis shows that while distal enhancers can be aberrantly silenced or activated in cancer (Figure 7), the data indicate that the process may be distinct from promoters and insulators. Foremost, our data suggest that the enhancer-associated H3K4me1 mark may not be mutually exclusive with DNA methylation, unlike H3K4me3, and that a change in nucleosome occupancy (gain or loss) may be sufficient to trigger epigenetic changes at enhancers.

The study of two cancer cell line models allowed a detailed and methodical investigation showing reconfiguration of nucleosome depleted regions at distal regulatory elements accompanies epigenetic changes that occur at enhancers and insulators. It is interesting to consider that the molecular changes we have observed in this analysis of breast and prostate cancer cells may explain some of the gross

alterations that occur in tumors and are visible to pathologists. Taken together, our findings suggest that disruption of the local nuclear architecture at enhancers and insulators offers a novel role for epigenetic reprogramming distant from promoters in cancer cells.

METHODS

Cell Culture

Normal human prostate epithelial cells (PrEC), a prostate cancer cell line (PC3) and a breast cancer cell line (MCF7) were obtained from the American Type Culture Collection (ATCC). Normal human mammary epithelial cells (HMEC) were obtained from Invitrogen. All cell lines were cultured under recommended conditions at 37°C and 5% CO₂.

Nucleosome Occupancy and Methylation Assay (NOMe Assay)

NOMe assays were conducted as described previously (Taberlay et al. 2011). Briefly, cells were trypsinized and centrifuged for 3 mins at 500g, then washed in ice-cold PBS and resuspended in 1mL ice-cold Nuclei Buffer (10mM Tris, pH 7.4, 10mM NaCl, 3mM MgCl₂, 0.1mM EDTA and 0.5% NP-40, plus protease inhibitors) per 5x10⁶ cells and incubated on ice for 5 min. Nuclei were recovered by centrifugation at 900g for 3 min and washed in Nuclei Wash Buffer (10mM Tris, pH 7.4, 10mM NaCl, 3mM MgCl₂ and 0.1mM EDTA containing protease inhibitors). Freshly prepared nuclei (2x10⁵ cells) were resuspended in 1X *M.CviPI* reaction buffer (NEB), then treated with 150U of *M.CviPI* (NEB; 50,000U/mL) in 15μL 10X reaction buffer, 45μL 1M sucrose and 0.75μL SAM in a volume of 150uL. Reactions were quenched by the addition of an equal volume of Stop Solution (20nM Tris-HCl [pH 7.9], 600mM NaCl, 1% SDS, 10mM EDTA, 400μg/ml Proteinase K) and incubated at 55°C overnight. DNA was purified by phenol/chloroform extraction and ethanol precipitation. Bisulfite conversion was performed using the EpiTect Bisulfite Kit (Qiagen). For analyses of individual genomic regions of interest, PCR amplicons were

cloned using the TOPO TA Kit (Invitrogen) and then sequenced. Oligonucleotides used for enhancer epigenetic silencing were:

Enh14For 5'- TATTTTATTATTAGGAATTTTGTAATTTTTTTAAG-3'

Enh14Rev 5'- AACCTCTACTTTATTTAATAATTTCTTCA -3'

Nucleosome Occupancy and DNA methylation Sequencing (NOMe-seq)

Libraries for genome-wide NOMe-seq analyses were prepared using the established protocols of the USC Epigenome Center. Briefly, genomic DNA (2µg) was sonicated using a Covaris instrument to an average molecular weight of 150 bp. Achievement of the desired size range was verified by Bioanalyzer analysis (Agilent Technologies). Fragmented DNA was repaired to generate blunt ends using the END-It kit (Epicentre Biotechnologies) according to manufacturers' instructions. Following incubation, the treated DNA was purified using AMPure XP beads (Agencourt). Magnetic beads were employed for all nucleic acid purifications in subsequent steps. Following end repair, A-tailing was performed using the dA-tailing module according to manufacturer's instructions (New England Biolabs). Adapters with a 3' 'T' overhangs were then ligated to the end-modified DNA. Modified Illumina paired-end (PE) adapters were used. Ligation was carried out using ultrapure, rapid T4 ligase (Enzymatics) according to manufacturer's instructions. The final product was then purified with magnetic beads to yield an adapter-ligation mix. Prior to bisulfite conversion, bacteriophage lambda DNA that had been through the same library preparation protocol described above to generate adapter-ligation mixes was combined with the genomic sample adapter ligation mix at 0.5% w/w. Adapter-ligation mixes were then bisulfite converted using the Zymo DNA Methylation Gold kit (Zymo Research) according to the manufacturers' recommendations. Final

modified product was purified by magnetic beads and eluted in a final volume of 20 μ L. Amplification of one-half the adapter-ligated library was performed using Kapa HiFi-U Ready Mix under the following conditions: 98°C for 2 min, followed by 4 cycles of 98°C for 30 s then 65°C for 15 s and 72°C for 1 min with a final extension for 10 min in 50 μ L total reaction volume. The final library product was examined on the Agilent Bioanalyzer then quantified using the Kapa Biosystems Library Quantification kit according to manufacturer's instructions. Optimal concentrations to get the right cluster density were determined empirically. Resulting libraries were plated using the Illumina cBot and run on the Illumina HiSeq 2000 platform configured for 100bp paired-end reads according to manufacturer's instructions. Read alignment was performed by the USC Epigenome Center (University of Southern California, Los Angeles) and analysis was performed using a modification of our existing bisulphite-sequencing pipeline (Statham et al. 2012) and custom scripts. QC data is provided in Supplementary Table 2. As described previously (Kelly et al. 2012), it is necessary to exclude GCG and CCG dinucleotides prior to analyses. This is particularly important since cytosine methylation cannot be designated endogenous (CpG) or enzymatic (GpC; from *M.CviPI* treatment) within GCG sites, while CCG can be the site of spurious methyltransferase activity (Kelly et al. 2012). GCGs equate to less than 0.24% of the genome and make up only 5.6% of all GpC sites leaving 94.4% of GpC sites intact for interrogation. On average there is 1 GpC dinucleotide every 20bp (average, 5 GpC sites per 100bp). Specifically, 78% of the genome has from 1-6 GpC sites per 100bp; Figure S6A) and detected NDRs ($-\log_{10}$ p-value ≥ 15 in any of the four samples) peak at 6-13 GpC sites (Figure S6B). In addition ~93% of GCGs contain a GCH within 20 bp (~46% being within 5 bp) (Kelly et al. 2012). Some depletion of GC-rich regions occurs in our data (Figure S6C), as expected

(Ross et al. 2013); however, GpC-rich regions (eg. >10 GpC sites per 100bp bin) account for only 2.06% of the human genome (Figure S6A) and therefore has minimal effect on our analysis.

Detection of Nucleosome Depleted Regions (NDRs)

The number of C and T nucleotides sequenced were counted at each GCH site in the genome, and sites with less than 5x coverage in all samples were discarded from further analysis. C and T counts were summed in 100bp windows at 20bp spacing and tested for difference to the genome background using the chi-squared test. Significant windows were scored at increasing p-value cutoffs (from $-5\log_{10}$ to $-20\log_{10}$), overlapped and only retained as NDRs if they were a minimum of 140bp in size.

Chromatin Immunoprecipitation (ChIP)

ChIP assays were performed as previously (Oakford et al. 2010; Taberlay et al. 2011). Briefly, nuclei were purified (described above for NOMe-seq) after formaldehyde crosslinking, collected and resuspended in SDS Lysis buffer before sonication. Antibodies (10 μ g) used for ChIP experiments included H3K4me1 (#39298, Active Motif), H3K4me3 (#39160, Active Motif), H3K27me3 (#39155, Active Motif), H3K27ac (#39297, Active Motif), RNA Polymerase II (#ab817, Abcam), CTCF (#07-729) and CD8 (#sc-32812, Santa Cruz). Libraries for ChIP-seq were prepared by the USC Epigenome Center following Illumina protocols. The resulting libraries were sequenced on the Illumina HiSeq 2000 platform configured for 50bp single-end reads. Public data was downloaded from ENCODE (Supplementary Table 1; (Meyer et al. 2011; Thurman et al. 2012; Wang et al. 2012a; Gertz et al. 2013)). Bowtie (Langmead et al. 2009) was used to align ChIP-seq reads to hg19 as previously described (Bert et

al. 2013) allowing up to 3 mismatches, discarding reads mapping to multiple positions in the genome and removing clonal reads. ChromHMM (Ernst and Kellis 2012) was applied to the chromatin modification, CTCF and RNA polymerase II aligned reads to simultaneously partition the genome of all four cell lines into 15 chromatin states. Redundant states were then collapsed into 9 distinct states and manually annotated by comparison to the published ChromHMM model for HMEC cells (Ernst et al. 2011). Transcription factor peaks were called using HOMER version 4.1 (Heinz et al. 2010) under the default parameters. DNase I peaks were called using Hotspot version 3 (John et al. 2011) with an FDR cutoff of 0.01.

DATA ACCESS

Raw and processed NOME-seq and ChIP-seq data has been deposited in the Gene Expression Omnibus (accession number GSE57498).

ACKNOWLEDGEMENTS

We thank members of the Clark and Jones Laboratories for helpful discussions and careful reading of the manuscript. SJC is a NHMRC Senior Principal Research Fellow; PCT is a Cancer Institute NSW Early Career Development Fellow. This work was further supported by Cure Cancer Australia Foundation Project Grant #1070613 and a NHMRC Project Grant #1051757.

DISCLOSURE DECLARATION

NOMe-seq has been commercialized by Active Motif. T.K.K is now an employee of Active Motif, Inc. but was not at the time of completion of this work. The authors have no other relevant affiliations, financial involvement or conflicts of interest with any organisation other than those stated.

FIGURE LEGENDS

Figure 1: NOMe-seq reveals that patterns of nucleosome occupancy are largely unique to individual cell types. A schematic representation of NOMe-seq methodology and the bioinformatics pipeline implemented for the detection of NDRs independent of other genomic features (eg. TSS, DNase hypersensitivity sites; see Methods).

Figure 2: NDRs are enriched at enhancers in normal epithelial cells, and change to predominantly overlap CTCF in cancer cells. (A-B) ChIP assays were performed with antibodies detecting RNA Pol II-P, H3K4me3, H3K27ac, H3K4me1 and H3K27me3 on chromatin from HMEC, MCF7, PrEC and PC3 cells. ChIP-seq data for CTCF was generated (PC3) or obtained from ENCODE. ChIP-seq chromatin states were classified using the ChromHMM Hidden Markov Model (Ernst and Kellis 2012). Data are presented as percent of NDRs detected by NOMe-seq in each of nine chromatin states (heterochromatin, repressed, transcribed, enhancer, enhancer + CTCF, CTCF, promoter + CTCF, promoter, promoter_poised) (y-axis) for HMEC (pink bars), MCF7 (purple bars), PrEC (blue bars) and PC3 (green bars). (C-D) As for (A-B) using DNase I hypersensitivity data as the measure of accessibility (% NDRs in each class).

Figure 3: Epigenetic silencing of enhancers and insulators, in addition to promoters, in cancer cells. (A) Promoters (broadly, nucleosome depleted and H3K4me3 enriched) were defined in HMEC cells. These exact genomic regions were subject to ChromHMM classification in MCF7 cells and the extent and type of epigenetic silencing (nucleosome acquisition, loss of active epigenetic marks and gain

of DNA methylation) was determined at promoters (left panel, teal). Similarly, insulators (broadly, nucleosome depleted and CTCF; center panel, purple) and enhancers (broadly, nucleosome depleted and H3K4me1 enriched; right panel, blue) were defined in HMEC cells and these exact genomic regions were subject to ChromHMM classification in MCF7 cells and the extent and type of epigenetic silencing was determined at insulators and enhancers. (B) As in (A), for PrEC and PC3. (C) Screenshots showing epigenetic reprogramming at a promoter (left panel), a CTCF site (middle panel) and an enhancer (right panel). For ChIP-seq tracks, orange represents CTCF; pink, H3K27ac; purple, H3K4me1; green, H3K4me3 and red, H3K27me3. (D) Validation of enhancer epigenetic silencing. Representative colours for ChIP-seq tracks are as shown in (C). All four cell types were treated with *M.CviPI* GpC methyltransferase and subjected to bisulfite conversion and cloning. Horizontal lines represent individual enhancers. Circles represent GpC dinucleotides (white, unmethylated and inaccessible to *M.CviPI*; teal, methylated and accessible to *M.CviPI*). Pink bars represent sites associated with nucleosomes. Regions accessible to *M.CviPI* (teal) indicate NDRs. Diagram is drawn to scale.

Figure 4: Epigenetic activation of enhancers and insulators, in addition to promoters, in cancer cells. (A) Promoters (broadly, nucleosome depleted and H3K4me3 enriched) were defined in MCF7 cells. These exact genomic regions were subject to ChromHMM classification in HMEC cells and the extent and type of epigenetic activation (nucleosome loss, acquisition of active epigenetic marks and reduced DNA methylation) was determined at promoters (left panel, teal). Similarly, for insulators (broadly, nucleosome depleted and CTCF; center panel, purple) and

enhancers (broadly, nucleosome depleted and H3K4me1 enriched; right panel, blue).

(B) as in (A), for PrEC and PC3 cells.

Figure 5: NOMe-seq reveals that nucleosomes are organized throughout the genome, at Nucleosome Depleted Regions (NDRs) and *facultative*NDRs (fNDRs).

(A-F) Each genomic feature (NDR, DNase hypersensitivity or CTCF binding site) was called as “present” (solid lines) or “absent” (broken lines) in HMEC (upper panels) or MCF7 (lower panels). Nucleosome occupancy and endogenous methylation (CpG) were mapped +/- 1000bp from the center (“0”) of each NDR, DNase I hypersensitive site or CTCF site. (A,B) NOMe-seq demonstrates that nucleosomes are present on either side of a NDR (solid teal lines) and across the genome, even at less pronounced or absent NDRs (broken teal line). DNA methylation is phased alongside nucleosomes (black lines) regardless of nucleosome occupancy; however, DNA methylation is typically low (~20%) within NDRs (solid black line). (C, D). DNase I hypersensitive sites are characterized by organized nucleosomes (solid teal lines) while DNA methylation (solid black lines) is low at the center then phased and increasing with distance from the center of the DNase I site (solid black line) similar to patterns observed around NDRs. (E, F). CTCF is associated with nucleosome patterns (solid teal lines) and phased peaks of DNA methylation (solid black lines) between the nucleosomes. In the absence of CTCF, nucleosomes are not phased (broken teal lines) and a continuous high level of DNA methylation is observed (~60%).

Figure 6: CTCF, RAD21 and REST are associated with patterns of nucleosome organization. (A-C) Nucleosome occupancy (teal lines) and endogenous methylation

(CpG; black lines) were mapped +/- 1000bp from the center ("0") of each transcription factor binding site and were determined to be "strongly phased" (CTCF, REST and RAD21, upper panel) in MCF7 cells.

Figure 7: Model depicting mechanisms by which the distal regulatory architecture provides an additional layer of epigenetic plasticity in cancer.

Enhancers preferentially undergo epigenetic silencing, which is exemplified by collapse of a NDR (nucleosome gain), DNA hypermethylation and loss of the H3K4me1 mark. Epigenetic switching can also encompass aberrant epigenetic activation of enhancers. Here, an atypical NDR is formed due to nucleosome loss and the enhancer becomes marked by H3K4me1, which is accompanied by DNA hypomethylation events in the cancer cells. Our data indicate that many more enhancers are abnormally silenced than activated, consistent with a hypothesis that it is more difficult to activate an already silenced regulatory element than it is to silence an already active or poised regulatory element. These findings also support permanently silenced state in cancer. Thus, our data demonstrates that the global epigenetic landscape is dynamically altered at key regulatory regions outside of promoters. NDR, nucleosome depleted region; small white circle, unmethylated CpG site; small black circle, methylated CpG site; large circle, nucleosome; 4me1 (green), H3K4me1.

SUPPLEMENTAL FIGURE LEGENDS

Supplementary Figure 1: NOMe-seq reveals that patterns of nucleosome occupancy are largely unique to individual cell types. (A-B) NDRs were called at increasing $-\log_{10}$ p-values (5-20) in all cell types. The number of NDRs (bars, left y-axis) and the distribution (filled circles, right y-axis) was determined at each cutoff. The $-\log_{10}$ p-value of 15 was used for all subsequent analyses. (C) NDRs were designated unique, or common to two, three or all four cell lines analyzed. (D) NDRs remain cell-type specific, irrespective of increasing p-value cutoffs. The percentage of NDRs is expressed as a proportion of the total at each cutoff. (E) Very few NDRs (<4%) are found in all four cell types, while the majority (~65%) are found in only one of the four cell types analyzed. HMEC, pink. MCF7, purple. Nucleosome depleted region (NDR), teal.

Supplementary Figure 2: Methods to detect accessible regions genome-wide have a different preference to GpC density. (A) Overlap between the number of accessible regions determined by DNase-seq and FAIRE-seq as well as NDRs detected by NOMe-Seq are plotted as a Venn diagram. (B-D) GpC density is plotted as the percent of bins overlapping (B) DNase I and (C) FAIRE accessible regions and (D) NOMe-seq NDRs.

Supplementary Figure 3: Validation of ChromHMM classification of the epigenomes of cell lines. (A) Heatmaps of average ChIP-seq enrichment signal observed for each mark in the 9 learned ChromHMM states. (B) Each ChromHMM state was integrated with RNA-seq data. The data are expressed as the mean RNA-seq signal (y-axis) on the positive strand (scale 0 to 15; solid line) and on the negative

strand (scale 0 to 15; dashed line). The distance from the center of the ChromHMM state is shown (x-axis; bp). NDR, nucleosome depleted region; bp, base pair.

Supplementary Figure 4: NOMe-seq reveals that nucleosomes are organized throughout the genome, at Nucleosome Depleted Regions (NDRs) and facultativeNDRs (fNDRs). (A-F) Each genomic feature (NDR, CTCF binding site or DNase I hypersensitivity site) was called as “present” (solid lines) or “absent” (broken lines) in PrEC (upper panels) or PC3 (lower panels). Nucleosome occupancy and endogenous methylation (CpG) were mapped +/- 1000bp from the center (“0”) of each NDR, DNase I hypersensitive site and CTCF site. (A, B) NOMe-seq demonstrates that nucleosomes are organized on either side of a NDR (solid teal lines) and across the genome, even at less pronounced or absent NDRs (broken teal line). DNA methylation is phased alongside nucleosomes (black lines) regardless of nucleosome occupancy; however, DNA methylation is typically low (~20%) within NDRs (solid black line). (C, D). DNase I hypersensitive sites are characterized by organized nucleosomes (solid teal lines) while DNA methylation (solid black lines) is low at the center then phased and increasing with distance from the center of the DNase I site (solid black line) similar to patterns observed around NDRs. (E, F). CTCF is associated with nucleosome patterns (solid teal lines) and phased peaks of DNA methylation (solid black lines) between the nucleosomes. In the absence of CTCF, nucleosomes are not phased (broken teal lines) and a continuous high level of DNA methylation is observed (~60%). (G-J) Nucleosome occupancy and endogenous methylation (CpG) were mapped +/- 1000bp from the center (“0”) of each NDR in all four cells types at a threshold of $-\log_{10} p\text{-value} = 3$ to demonstrate that phasing still occurs when NDRs are completely absent (broken teal lines).

Supplementary Figure 5: Multiple transcription factors are associated weakly phased or unorganized nucleosomes. (A-B) Nucleosome occupancy (teal lines) and endogenous methylation (CpG; black lines) were mapped +/- 1000bp from the center ("0") of each transcription factor binding site and were determined to be (A) "weakly phased" or (B) unphased in MCF7 cells.

Supplementary Figure 6: NOMe-seq metrics associated with genomic GpC density. (A) Distribution of GpC density within 100bp bins, (B) NDRs detected relative to GpC density and (C) sequencing coverage relative to GpC density were calculated.

REFERENCES

- Akhtar-Zaidi B, Cowper-Sal-lari R, Corradin O, Saiakhova A, Bartels CF, Balasubramanian D, Myeroff L, Lutterbaugh J, Jarrar A, Kalady MF et al. 2012. Epigenomic enhancer profiling defines a signature of colon cancer. *Science* **336**(6082): 736-739.
- Almeida A, Kokalj-Vokac N, Lefrancois D, Viegas-Pequignot E, Jeanpierre M, Dutrillaux B, Malfoy B. 1993. Hypomethylation of classical satellite DNA and chromosome instability in lymphoblastoid cell lines. *Hum Genet* **91**(6): 538-546.
- Andreu-Vieyra C, Lai J, Berman BP, Frenkel B, Jia L, Jones PA, Coetzee GA. 2011. Dynamic nucleosome-depleted regions at androgen receptor enhancers in the absence of ligand in prostate cancer cells. *Mol Cell Biol* **31**(23): 4648-4662.
- Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* **129**(4): 823-837.
- Baylin SB, Jones PA. 2011. A decade of exploring the cancer epigenome - biological and translational implications. *Nature reviews Cancer* **11**(10): 726-734.
- Berman BP, Weisenberger DJ, Aman JF, Hinoue T, Ramjan Z, Liu Y, Noushmehr H, Lange CP, van Dijk CM, Tollenaar RA et al. 2012. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nature genetics* **44**(1): 40-46.
- Bernstein BE, Liu CL, Humphrey EL, Perlstein EO, Schreiber SL. 2004. Global nucleosome occupancy in yeast. *Genome Biol* **5**(9): R62.
- Bert SA, Robinson MD, Strbenac D, Statham AL, Song JZ, Hulf T, Sutherland RL, Coolen MW, Stirzaker C, Clark SJ. 2013. Regional activation of the cancer genome by long-range epigenetic remodeling. *Cancer cell* **23**(1): 9-22.
- Collings CK, Waddell PJ, Anderson JN. 2013. Effects of DNA methylation on nucleosome stability. *Nucleic Acids Res* **41**(5): 2918-2931.
- Coolen MW, Stirzaker C, Song JZ, Statham AL, Kassir Z, Moreno CS, Young AN, Varma V, Speed TP, Cowley M et al. 2010. Consolidation of the cancer genome into domains of repressive chromatin by long-range epigenetic silencing (LRES) reduces transcriptional plasticity. *Nature cell biology* **12**(3): 235-246.
- Ernst J, Kellis M. 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature biotechnology* **28**(8): 817-825.
- . 2012. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**(3): 215-216.
- Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M et al. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**(7345): 43-49.
- Felle M, Hoffmeister H, Rothhammer J, Fuchs A, Exler JH, Langst G. 2011. Nucleosomes protect DNA from DNA methylation in vivo and in vitro. *Nucleic Acids Res* **39**(16): 6956-6969.
- Field Y, Fondufe-Mittendorf Y, Moore IK, Mieczkowski P, Kaplan N, Lubling Y, Lieb JD, Widom J, Segal E. 2009. Gene expression divergence in yeast is coupled to evolution of DNA-encoded nucleosome organization. *Nature genetics* **41**(4): 438-445.

- Field Y, Kaplan N, Fondufe-Mittendorf Y, Moore IK, Sharon E, Lubling Y, Widom J, Segal E. 2008. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS computational biology* **4**(11): e1000216.
- Gal-Yam EN, Egger G, Iniguez L, Holster H, Einarsson S, Zhang X, Lin JC, Liang G, Jones PA, Tanay A. 2008. Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. *Proceedings of the National Academy of Sciences of the United States of America* **105**(35): 12979-12984.
- Gertz J, Savic D, Varley KE, Partridge EC, Safi A, Jain P, Cooper GM, Reddy TE, Crawford GE, Myers RM. 2013. Distinct properties of cell-type-specific and shared transcription factor binding sites. *Molecular cell* **52**(1): 25-36.
- Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res* **17**(6): 877-885.
- Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW et al. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**(7243): 108-112.
- Heintzman ND, Ren B. 2007. The gateway to transcription: identifying, characterizing and understanding promoters in the eukaryotic genome. *Cell Mol Life Sci* **64**(4): 386-400.
- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**(4): 576-589.
- Hinshelwood RA, Melki JR, Huschtscha LI, Paul C, Song JZ, Stirzaker C, Reddel RR, Clark SJ. 2009. Aberrant de novo methylation of the p16INK4A CpG island is initiated post gene silencing in association with chromatin remodelling and mimics nucleosome positioning. *Human molecular genetics* **18**(16): 3098-3109.
- Hogan GJ, Lee CK, Lieb JD. 2006. Cell cycle-specified fluctuation of nucleosome occupancy at gene promoters. *PLoS genetics* **2**(9): e158.
- Hoose SA, Kladde MP. 2006. DNA methyltransferase probing of DNA-protein interactions. *Methods Mol Biol* **338**: 225-244.
- Hughes AL, Jin Y, Rando OJ, Struhl K. 2012. A functional evolutionary approach to identify determinants of nucleosome positioning: a unifying model for establishing the genome-wide pattern. *Molecular cell* **48**(1): 5-15.
- Ioshikhes IP, Albert I, Zanton SJ, Pugh BF. 2006. Nucleosome positions predicted through comparative genomics. *Nature genetics* **38**(10): 1210-1215.
- Jeong S, Liang G, Sharma S, Lin JC, Choi SH, Han H, Yoo CB, Egger G, Yang AS, Jones PA. 2009. Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA. *Mol Cell Biol* **29**(19): 5366-5376.
- Jessen WJ, Dhasarathy A, Hoose SA, Carvin CD, Risinger AL, Kladde MP. 2004. Mapping chromatin structure in vivo using DNA methyltransferases. *Methods* **33**(1): 68-80.
- Jessen WJ, Hoose SA, Kilgore JA, Kladde MP. 2006. Active PHO5 chromatin encompasses variable numbers of nucleosomes at individual promoters. *Nat Struct Mol Biol* **13**(3): 256-263.

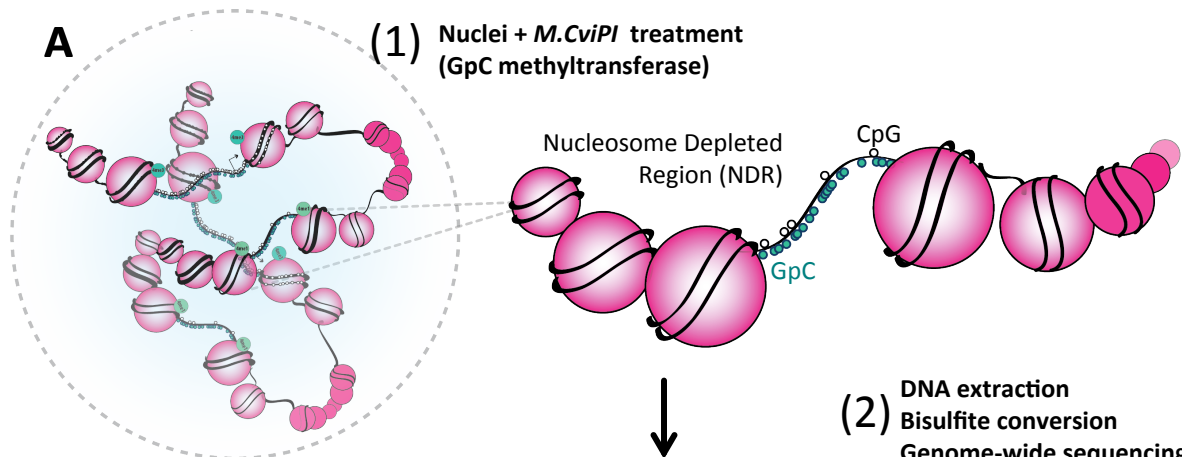
- Ji W, Hernandez R, Zhang XY, Qu GZ, Frady A, Varela M, Ehrlich M. 1997. DNA demethylation and pericentromeric rearrangements of chromosome 1. *Mutat Res* **379**(1): 33-41.
- John S, Sabo PJ, Thurman RE, Sung MH, Biddie SC, Johnson TA, Hager GL, Stamatoyannopoulos JA. 2011. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nature genetics* **43**(3): 264-268.
- Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* **13**(7): 484-492.
- Jones PA, Baylin SB. 2007. The epigenomics of cancer. *Cell* **128**(4): 683-692.
- Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J et al. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**(7236): 362-366.
- Kelly TK, Jones PA. 2011. Role of nucleosomes in mitotic bookmarking. *Cell cycle* **10**(3): 370-371.
- Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, Jones PA. 2012. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res* **22**(12): 2497-2506.
- Kelly TK, Miranda TB, Liang G, Berman BP, Lin JC, Tanay A, Jones PA. 2010. H2A.Z maintenance during mitosis reveals nucleosome shifting on mitotically silenced genes. *Molecular cell* **39**(6): 901-911.
- Kilgore JA, Hoose SA, Gustafson TL, Porter W, Kladde MP. 2007. Single-molecule and population probing of chromatin structure using DNA methyltransferases. *Methods* **41**(3): 320-332.
- Kim A, Song SH, Brand M, Dean A. 2007. Nucleosome and transcription activator antagonism at human beta-globin locus control region DNase I hypersensitive sites. *Nucleic Acids Res* **35**(17): 5831-5838.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**(3): R25.
- Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C. 2007. A high-resolution atlas of nucleosome occupancy in yeast. *Nature genetics* **39**(10): 1235-1244.
- Lin JC, Jeong S, Liang G, Takai D, Fatemi M, Tsai YC, Egger G, Gal-Yam EN, Jones PA. 2007. Role of nucleosomal occupancy in the epigenetic silencing of the MLH1 CpG island. *Cancer Cell* **12**(5): 432-444.
- Meyer KB, Maia AT, O'Reilly M, Ghousaini M, Prathalingam R, Porter-Gill P, Ambis S, Prokunina-Olsson L, Carroll J, Ponder BA. 2011. A functional variant at a prostate cancer predisposition locus at 8q24 is associated with PVT1 expression. *PLoS genetics* **7**(7): e1002165.
- Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, Thurman RE, John S, Sandstrom R, Johnson AK et al. 2012. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**(7414): 83-90.
- Oakford PC, James SR, Qadi A, West AC, Ray SN, Bert AG, Cockerill PN, Holloway AF. 2010. Transcriptional and epigenetic regulation of the GM-CSF promoter by RUNX1. *Leukemia research* **34**(9): 1203-1213.
- Prendergast GC, Lawe D, Ziff EB. 1991. Association of Myn, the murine homolog of max, with c-Myc stimulates methylation-sensitive DNA binding and ras cotransformation. *Cell* **65**(3): 395-407.

- Prendergast GC, Ziff EB. 1991. Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region. *Science* **251**(4990): 186-189.
- Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, Nusbaum C, Jaffe DB. 2013. Characterizing and measuring bias in sequence data. *Genome Biol* **14**(5): R51.
- Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K. 2008. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**(5): 887-898.
- Segal E, Fondufe-Mittendorf Y, Chen L, Thastrom A, Field Y, Moore IK, Wang JP, Widom J. 2006. A genomic code for nucleosome positioning. *Nature* **442**(7104): 772-778.
- Sharma S, De Carvalho DD, Jeong S, Jones PA, Liang G. 2011. Nucleosomes containing methylated DNA stabilize DNA methyltransferases 3A/3B and ensure faithful epigenetic inheritance. *PLoS genetics* **7**(2): e1001286.
- Shivaswamy S, Bhinge A, Zhao Y, Jones S, Hirst M, Iyer VR. 2008. Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS biology* **6**(3): e65.
- Song L, Zhang Z, Grasfeder LL, Boyle AP, Giresi PG, Lee BK, Sheffield NC, Graf S, Huss M, Keefe D et al. 2011. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res* **21**(10): 1757-1767.
- Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D et al. 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**(7378): 490-495.
- Statham AL, Robinson MD, Song JZ, Coolen MW, Stirzaker C, Clark SJ. 2012. Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. *Genome Res* **22**(6): 1120-1127.
- Struhl K, Segal E. 2013. Determinants of nucleosome positioning. *Nat Struct Mol Biol* **20**(3): 267-273.
- Taberlay PC, Jones PA. 2011. DNA methylation and cancer. *Progress in drug research Fortschritte der Arzneimittelforschung Progres des recherches pharmaceutiques* **67**: 1-23.
- Taberlay PC, Kelly TK, Liu CC, You JS, De Carvalho DD, Miranda TB, Zhou XJ, Liang G, Jones PA. 2011. Polycomb-repressed genes have permissive enhancers that initiate reprogramming. *Cell* **147**(6): 1283-1294.
- Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernet B et al. 2012. The accessible chromatin landscape of the human genome. *Nature* **489**(7414): 75-82.
- Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, Zeng K, Malek JA, Costa G, McKernan K et al. 2008. A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res* **18**(7): 1051-1063.
- Visel A, Bristow J, Pennacchio LA. 2007. Enhancer identification through comparative genomics. *Semin Cell Dev Biol* **18**(1): 140-152.
- Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, Lee K, Canfield T, Weaver M, Sandstrom R et al. 2012a. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res* **22**(9): 1680-1688.

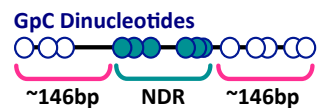
- Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, Greven MC, Pierce BG, Dong X, Kundaje A, Cheng Y et al. 2012b. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res* **22**(9): 1798-1812.
- Whitehouse I, Rando OJ, Delrow J, Tsukiyama T. 2007. Chromatin remodelling at promoters suppresses antisense transcription. *Nature* **450**(7172): 1031-1035.
- Wolff EM, Byun HM, Han HF, Sharma S, Nichols PW, Siegmund KD, Yang AS, Jones PA, Liang G. 2010. Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer. *PLoS genetics* **6**(4): e1000917.
- Xu M, Kladde MP, Van Etten JL, Simpson RT. 1998. Cloning, characterization and expression of the gene coding for a cytosine-5-DNA methyltransferase recognizing GpC. *Nucleic Acids Res* **26**(17): 3961-3966.
- You JS, De Carvalho DD, Dai C, Liu M, Pandiyan K, Zhou XJ, Liang G, Jones PA. 2013. SNF5 Is an Essential Executor of Epigenetic Regulation during Differentiation. *PLoS Genet* **9**(4): e1003459.
- You JS, Kelly TK, De Carvalho DD, Taberlay PC, Liang G, Jones PA. 2011. OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes. *Proc Natl Acad Sci U S A* **108**(35): 14497-14502.
- Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ. 2005. Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309**(5734): 626-630.

A

(1) Nuclei + *M.CviPI* treatment
(GpC methyltransferase)



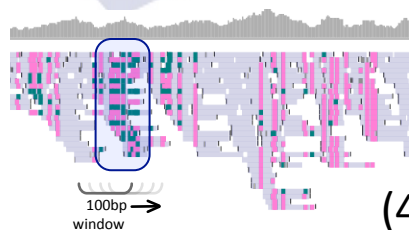
(2) DNA extraction
Bisulfite conversion
Genome-wide sequencing



(3) Sequence alignment &
Count GpC methylation

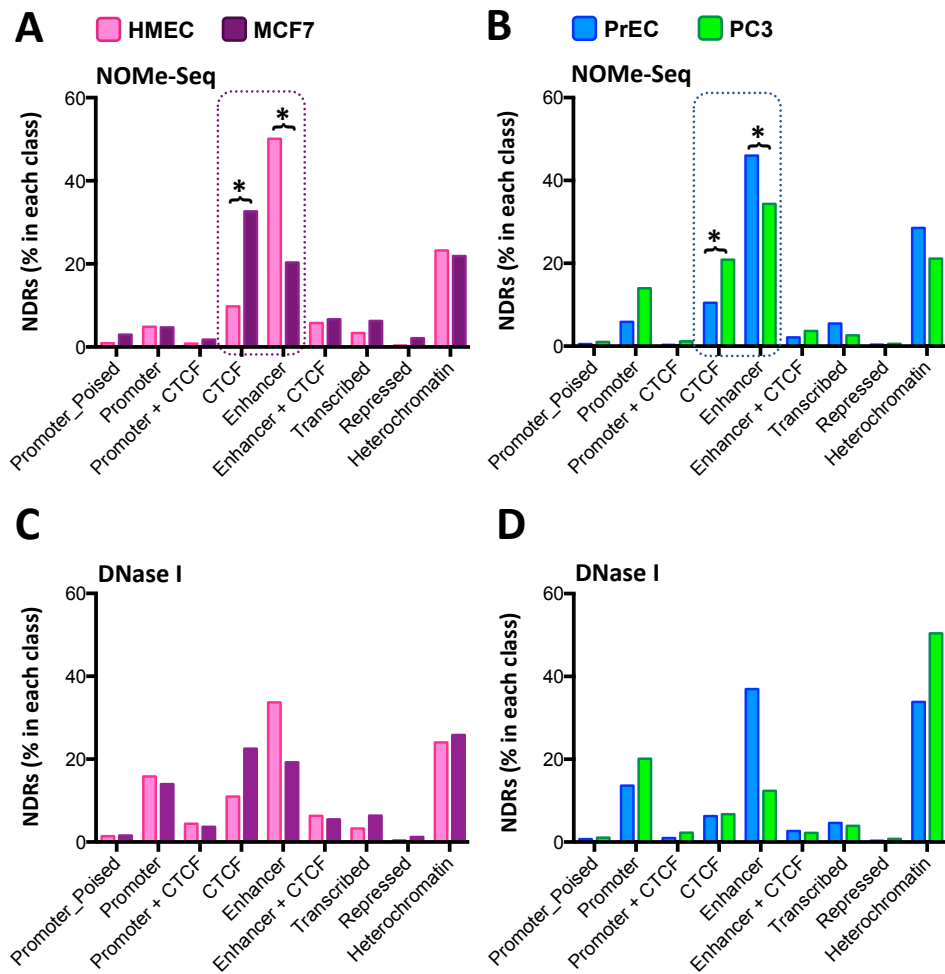


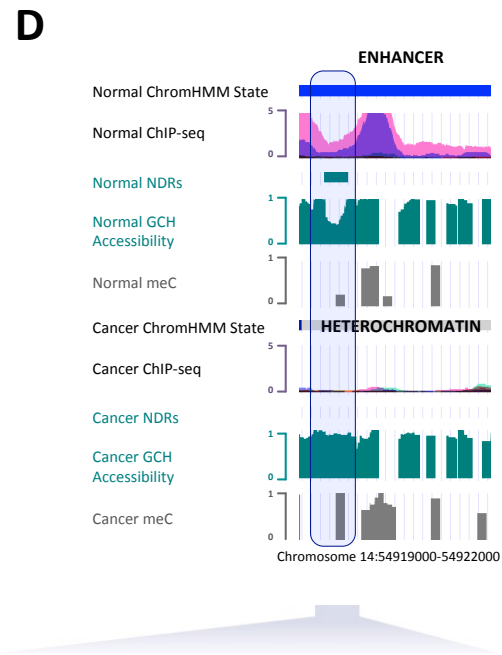
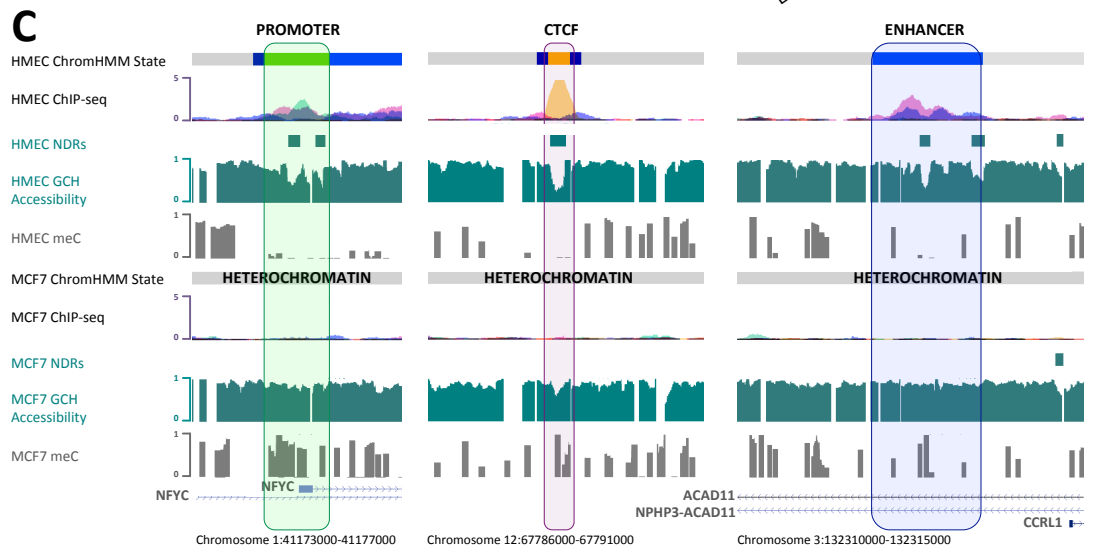
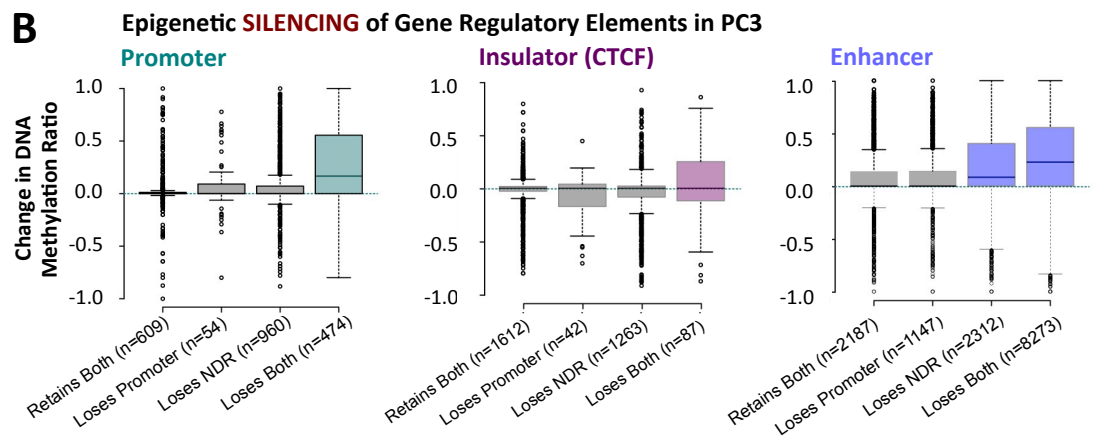
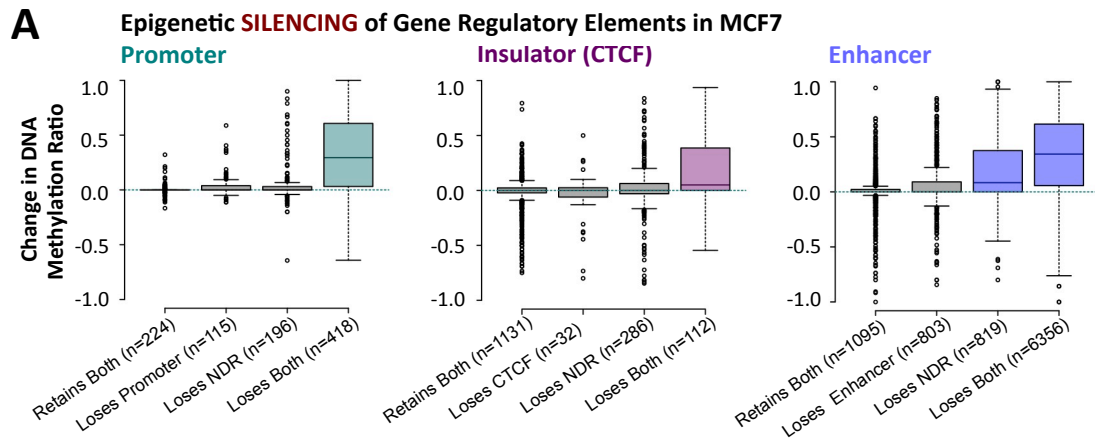
(4) Sliding window &
p-value calculation



(5) Call significant windows,
merge overlaps and filter by size
(retain >140bp)

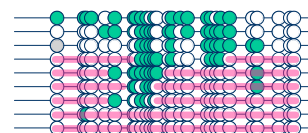




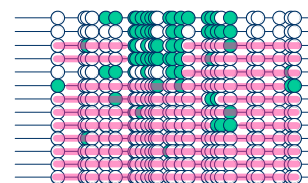


HMEC

Enhancer (Nuclei + MCviPI)
GpC Methylation (Accessibility)



PrEC

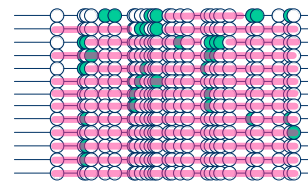


NORMAL

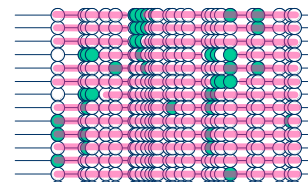
CANCER

MCF7

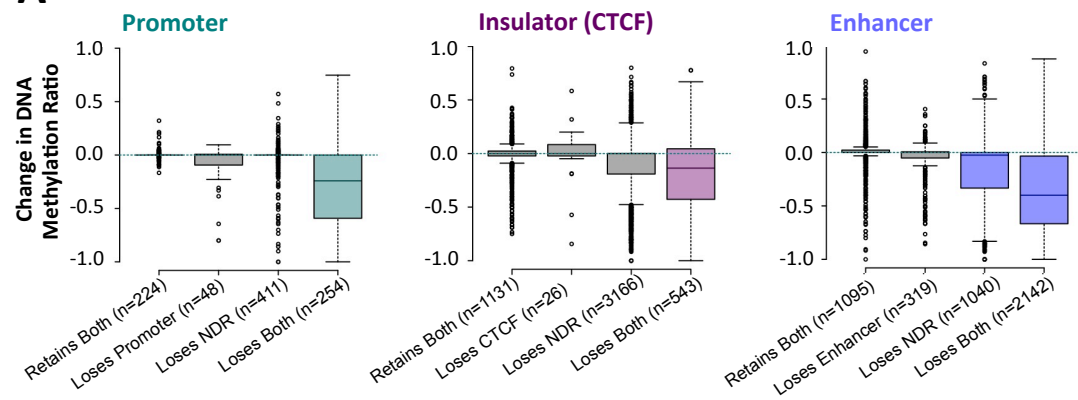
Enhancer (Nuclei + MCviPI)
GpC Methylation (Accessibility)



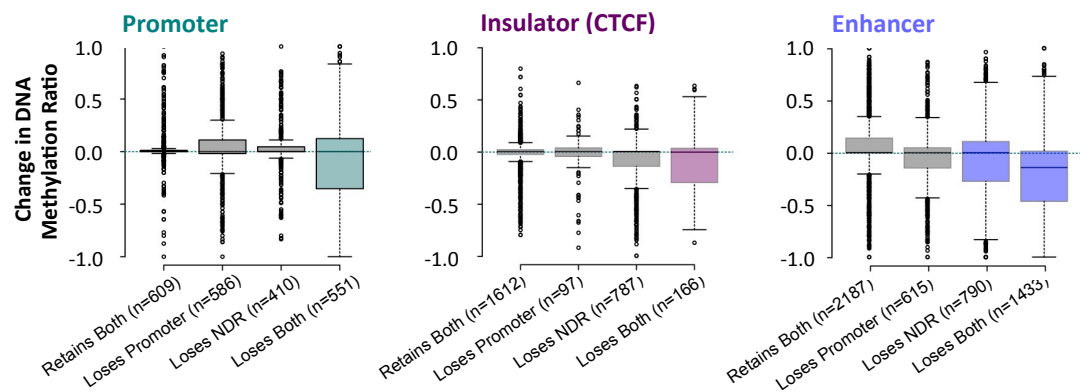
PC3



A Epigenetic ACTIVATION of Gene Regulatory Elements in MCF7

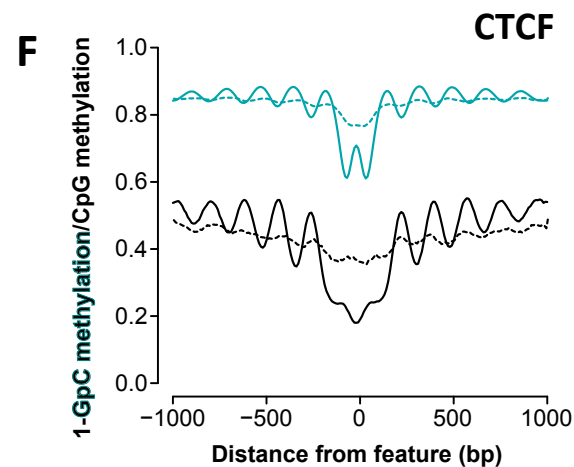
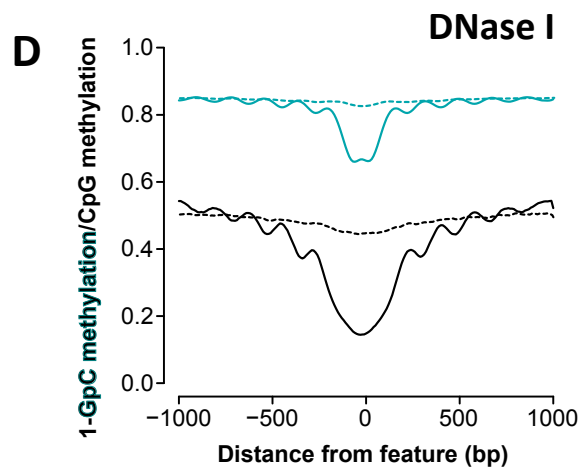
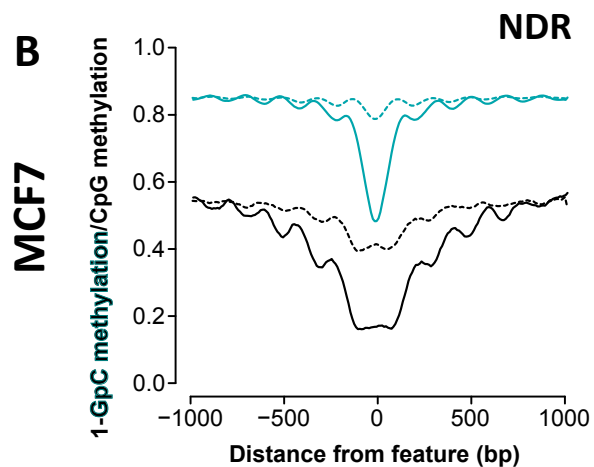
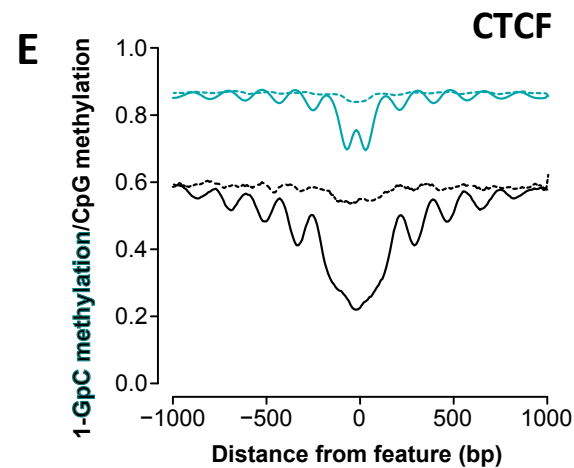
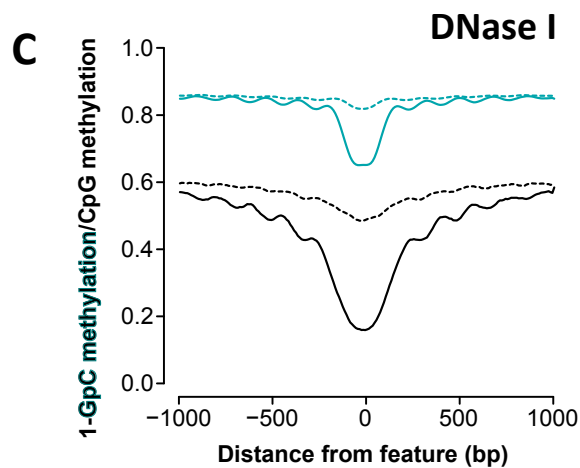
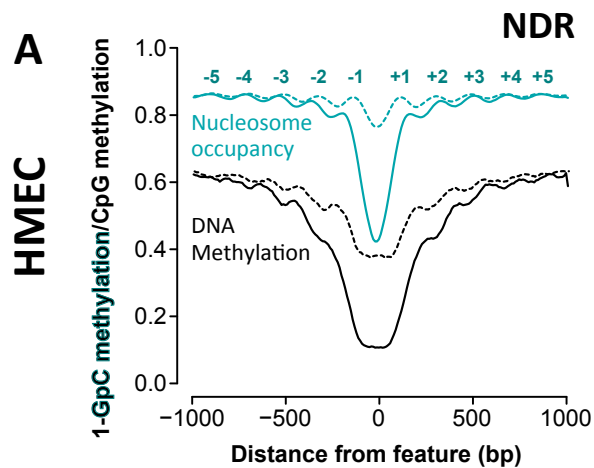


B Epigenetic ACTIVATION of Gene Regulatory Elements in PC3

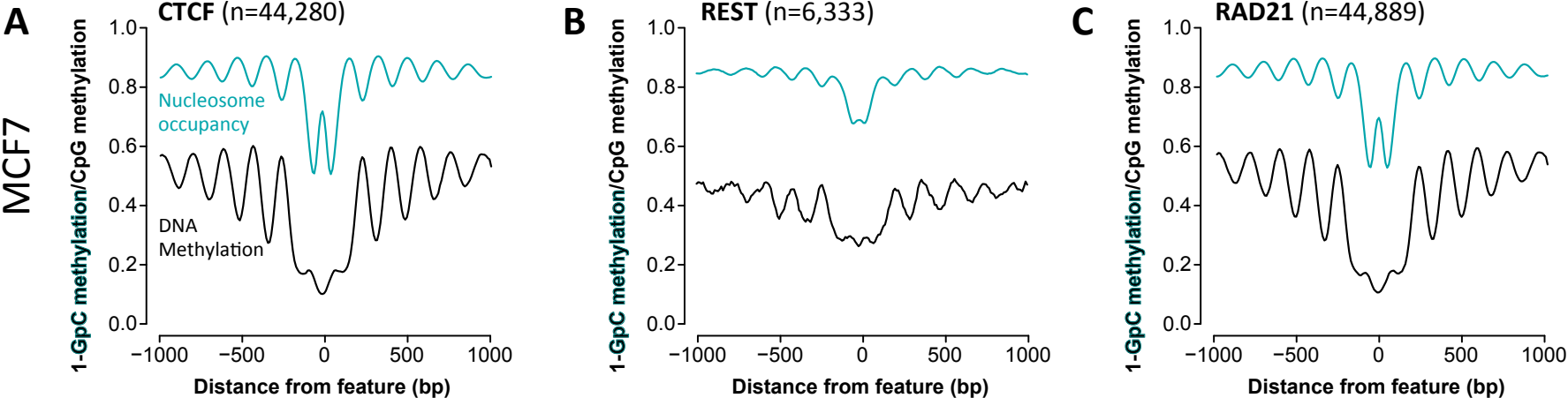


— Nucleosome positioning, feature present
- - Nucleosome positioning, feature absent (*facultative*)

— DNA methylation, feature present
- - DNA methylation, feature absent (*facultative*)



STRONGLY PHASED



**ENHANCER
EPIGENETIC
SILENCING**

**ENHANCER
EPIGENETIC
ACTIVATION**

