Published in final edited form as:

J Cell Physiol. 2017 June ; 232(6): 1295–1305. doi:10.1002/jcp.25596.

Identifying Nuclear Matrix–attached DNA across the Genome

Jason R. Dobson¹, Deli Hong², A. Rasim Barutcu¹, Hai Wu², Anthony N. Imbalzano¹, Jane B. Lian^{1,2}, Janet L. Stein^{1,2}, Andre J. van Wijnen³, Jeffrey A. Nickerson¹, and Gary S. Stein^{1,2,*}

¹Department of Cell and Developmental Biology, University of Massachusetts Medical School, Worcester, MA 01655

²Department of Biochemistry and University of Vermont Cancer Center, University of Vermont College of Medicine, Burlington, VT 05405

³Departments of Orthopedic Surgery and Biochemistry & Molecular Biology, Mayo Clinic, Rochester, MN 55905

Abstract

Experimental approaches to define the relationship between gene expression and nuclear matrix attachment regions (MARs) have given contrasting and method-specific results. We have developed a next generation sequencing strategy to identify MARs across the human genome (MAR-Seq). The method is based on crosslinking chromatin to its nuclear matrix attachment sites to minimize changes during biochemical processing. We used this method to compare nuclear matrix organization in MCF-10A mammary epithelial-like cells and MDA-MB-231 breast cancer cells and evaluated the results in the context of global gene expression (array analysis) and positional enrichment of gene-regulatory histone modifications (ChIP-Seq). In the normal-like cells, nuclear matrix–attached DNA was enriched in expressed genes, while in the breast cancer cells, it was enriched in non-expressed genes. In both cell lines, the chromatin modifications that mark transcriptional activation or repression were appropriately associated with gene expression. Using this new MAR-Seq approach, we provide the first genome-wide characterization of nuclear matrix attachment in mammalian cells and reveal that the nuclear matrix–associated genome is highly cell-context dependent.

Keywords

Genome structure; Nuclear matrix; Nuclear matrix–attached region; Nuclear matrix–attached DNA; Higher order nuclear structure

Correspondence to: Gary S. Stein, Department of Biochemistry and University of Vermont Cancer Center, University of Vermont College of Medicine, 89 Beaumont Avenue, Burlington, VT 05405; T: 802-656-4874; F: 802-656-2140; gary.stein@uvm.edu. Present address for Jason R. Dobson: Novartis Institutes for BioMedical Research, Oncology Research, Cambridge, MA 02139; Present address for A. Rasim Barutcu: Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138; Present address for Hai Wu: Yurogen Biosystems LCC, Worcester, MA 01605.

The authors have no conflicts of interest.

INTRODUCTION

In the mammalian cell, two meters of DNA are packaged in a nucleus that measures just 10 microns in diameter. This remarkable degree of compaction is accomplished by mechanisms that leave thousands of cell-type appropriate genes available for transcription (Dekker, 2014; Pombo and Dillon, 2015; Razin and Gavrilov, 2014). DNA is wrapped around nucleosomes and the resulting beads-on-a-string structure is condensed into 30 nm fibers that are further organized into chromatin loops ranging in size from 5 to 200 kb (Davie, 1995; Nelson et al., 1986; Razin et al., 1995). Such loops require structural constraints at their bases; they are formed by periodic attachments of the genome to the nuclear matrix (Benyajati and Worcel, 1976; Cook and Brazell, 1975; Cook et al., 1976; Vogelstein et al., 1980), a ribonucleoprotein (RNP) structural framework of the nucleus (Fey et al., 1986; He et al., 1990; Nickerson et al., 1997). In their now classic studies of mitotic chromosomes, Laemmli and colleagues observed a similar looping structure in which DNA is attached to a proteinaceous chromosomal scaffold (Paulson and Laemmli, 1977). During replication of the genome, stably anchored sites of DNA replication have been observed in living cells (Leonhardt et al., 2000). More recent genomics studies have revealed a higher-order nuclear organization in interphase cells that consists of chromosome territories, genomic compartments and topological associating domains (Gibcus and Dekker, 2013).

Some molecular details about chromatin-nuclear matrix attachment sites have been identified. Active genes have been localized to the nuclear matrix-anchored bases of chromatin loops as determined by fluorescence in situ hybridization (Gerdes et al., 1994) or indirect measurements (Maya-Mendoza and Aranda-Anzaldo, 2003; Maya-Mendoza et al., 2003). Genes can attach to the nuclear substructure through DNA sequences called matrix attachment regions (MARs) or scaffold attachment regions (SARs). These regions have 150–200 bp subregions that can become single stranded under torsional stress (Bode et al., 1992), consistent with AT enrichment (Flickinger, 1986). MARs are interspersed in the *Drosophila* genome at intervals of 26,000–112,000 bp (Mirkovitch et al., 1986)—the estimated sizes of chromatin loops in flies and mammals (Benyajati and Worcel, 1976; Jackson et al., 1990; Razin et al., 1995; Vogelstein et al., 1980). MAR binding to the nuclear matrix has been reported to both enhance and repress gene expression (Boulikas, 1995).

Using a method that includes high salt (2M NaCl) nuclear matrix isolation after nuclease digestion, more than 7,500 MARs have been recently identified in *Drosophila* embryos (Pathak et al., 2014). These sequences are highly AT-rich and predominantly located in noncoding regions of the genome; 53% are in introns. An earlier study also found that *Drosophila* MARs are enriched in AT-rich sequences, but presented evidence suggesting that the high-salt extraction causes sliding of DNA through matrix attachment sites, making it difficult to precisely localize MARs (Mirkovitch et al., 1984). These results greatly complicate analyses of the literature on MARs and the relationship of nuclear matrix attachment to gene expression. In recent practice, nuclear matrix attachments are called MARs when high-salt extractions are used in the preparations, while they are designated SARs (scaffold attachment regions) when lithium 3,5-diiodosalicylate extraction is substituted, as described by Mirkovitch et al. (Mirkovitch et al., 1984). A study comparing these methods for HeLa chromosomes 14 to 18 found that only half of all attachment sites

identified are common (Linnemann et al., 2009). SARS were more often 5' of a gene and were associated with the presence of a transcript. MARs were contained within a gene body and, compared to SARs, were more associated with silenced genes. Neither method identified random attachment sites in the genome. Both identified sites with a preferred positioning relative to genes and transcription, consistent with biological significance. The differences, however, suggest that either or both methodologies are flawed. There is clearly a need for a better method for identifying nuclear matrix/scaffold attachment sites of the genome in living cells and tissues, one that prevents sliding of DNA sequences through attachment sites.

To provide a method with reduced potential for artefactual rearrangements we have developed MAR-Seq, an optimized method for identifying nuclear matrix attachment sites in the genome after extensive crosslinking to prevent rearrangements (Nickerson et al., 1997). MAR-Seq uses a formaldehyde crosslinking protocol that is similar to the well-accepted methods typically used for chromatin immunoprecipitation and chromosome conformation capture experiments. It minimizes the possibility of artefactual chromatin attachment sites and captures native interactions of DNA with the nuclear matrix. We validated the approach by immunofluorescence analysis of nuclear bodies/structures both before and after isolation of the nuclear matrix and associated DNA. Using our new method, we observed cell type-specific patterns of nuclear matrix DNA association in non-tumorigenic MCF-10A mammary epithelial cells and metastatic, triple negative MDA-MB-231 breast cancer cells. Our results indicate that nuclear matrix-attached DNA is enriched in expressed genes in MCF-10A cells, and with non-expressed genes in MDA-MB-231 cells. The chromatin modifications associated with transcriptional activation (H3K4me3) or repression (H3K27me3 and H3K9me2) were found to be appropriately associated with expressed or non-expressed genes in both cell lines. Furthermore, the positions of matrix-attached regions in the MCF-10A genome are strongly correlated with open compartments identified using Hi-C analysis. These observations suggest that the nuclear matrix-associated organization of the genome is highly cell-context dependent. Taken together, our results bridge the gap between the biochemical and genomic views of the eukaryotic nucleus.

MATERIALS AND METHODS

Cell culture

MCF-10A cells, from a lab-maintained stock, were cultured in phenol-free DMEM/F12 media supplemented with EGF, insulin, cholera toxin and hydrocortisone and subcultured as described (Debnath et al., 2002). MDA-MB-231 cells, also from a lab-maintained stock, were grown and subcultured as previously described (Pratap et al., 2008). Cell lines are monitored for integrity by STR analysis and are routinely checked for mycoplasma contamination.

Immunofluorescence microscopy

Cells were grown on coverslips coated in 0.5% w/v gelatin as previously described prior to extraction (Ali et al., 2012; Ali et al., 2010). Fixed cells/nuclei/nuclear matrices were then incubated with primary antibodies (NPAT, Coilin, PML, SC-35, UBF and Pol II) prior to

incubation with Alexa Fluor-conjugated secondary antibodies as previously described (Ali et al., 2012; Ali et al., 2010; Ghule et al., 2009). Immunofluorescence was executed in parallel to preparation of nuclear matrices for genomic DNA extract and deep sequencing.

The following antibodies were used for immunofluorescence: polyclonal p220NPAT (rabbit polyclonal) (Ma et al., 2000; Zhao et al., 2000), coilin (pdelta) (mouse monoclonal, Santa Cruz Biotech), PML (H-238; rabbit polyclonal, Santa Cruz Biotechnology), mouse monoclonal antibody against SC35 (Sigma–Aldrich), RNA pol II (mouse monoclonal clone 8WG16, Covance, Princeton, NJ), and UBF (F-9, Santa Cruz Biotechnology).

Isolation of nuclear matrix-attached DNA and preparation for deep sequencing

Nuclear matrix bound DNA sequences were isolated using a modification of the method described previously (Nickerson et al., 1997). Confluent/nearly confluent cells were washed twice in cold PBS and incubated for 10 min on ice in cytoskeletal (CSK) extraction buffer (100 mM NaCl, 1 mM EGTA, 10 mM PIPES pH 6.8, 3 mM MgCl₂, 300 mM sucrose, 0.5% v/v Triton X-100). Cells were then washed twice with CSK wash buffer (100 mM NaCl, 1 mM EGTA, 10 mM PIPES pH 6.8, 3 mM MgCl₂, 300 mM sucrose, 1 mM EDTA) and incubated in CSK crosslinking buffer (100 mM NaCl, 1 mM EGTA, 10 mM PIPES pH 6.8, 3 mM MgCl₂, 300 mM sucrose, 1% w/v formaldehyde) on ice for 10 min with gentle agitation. Glycine solution (2.5 M) was added to halt crosslinking, after 10 min, the glycine solution was removed and plates were washed twice in CSK wash buffer. Next, cells were incubated for 1h at 30°C in nuclear matrix digestion buffer containing a cocktail of restriction enzymes (50 mM NaCl, 1 mM EGTA, 10 mM PIPES pH 6.8, 3 mM MgCl₂, 300 mM sucrose, 0.5% v/v Triton X-100, 600 U/mL MspI, 600 U/mL HaeIII, 600 U/mL RsaI, 600 U/mL AluI, 600 U/mL BamHI, 600 U/mL PvuII).

Plates were then placed on ice, washed twice with digestion wash buffer (50 mM NaCl, 11 mM EGTA, 10 mM PIPES pH 6.8, 3 mM MgCl₂, 300 mM sucrose, 10 mM EDTA) and incubated on ice for 10 min in matrix extraction buffer (50 mM NaCl, 11 mM EGTA, 10 mM PIPES pH 6.8, 3 mM MgCl₂, 300 mM sucrose, 10 mM EDTA, 250 mM ammonium sulfate) to further purify nuclear matrices and attached DNA. Matrix extraction buffer was aspirated off samples, they were washed twice more with digestion wash buffer, and then submerged in crosslink reversal buffer (100 mM NaHCO₃, 0.5% w/v SDS, 10 mM EDTA, 10 mM EGTA). Cells and buffer were scraped into a 50 mL centrifugal concentrator with a 10,000 molecular weight cut-off (Amicon), 10 mL additional crosslink reversal buffer was added and samples were centrifuged to concentrate to ~400 µL; this was repeated a second time. The preparation was transferred to a microcentrifuge tube and digested with 200 µg/mL Proteinase K at 65°C for 16h.

Finally DNA was isolated by extraction using phenol:chloroform:isoamyl alcohol (25:24:1) followed by chloroform:isoamyl alcohol (24:1) and RNA was digested by incubating with 200 µg/mL RNase A for 2h at 37°C. RNase was removed with a second phenol:chloroform:isoamyl alcohol/ chloroform:isoamyl alcohol extraction series and DNA was treated with 200 µg/mL Proteinase K for 2h at 55°C. Samples were then extracted a third time with phenol:chloroform:isoamyl alcohol, followed by chloroform:isoamyl alcohol, and precipitated overnight at –20°C in 100% ethanol.

The resulting MAR-Seq DNA was quantified, fragmented by sonication, and converted into libraries for deep sequencing on the Illumina Hi-Seq 2000 platform, following the Illumina TruSeq DNA Sample Preparation v2 Guide. Two replicates were processed, normalized and the data combined for analyses.

ChIP-Seq

ChIP was performed on subconfluent cells as previously described (Lee et al., 2006) with the following modification: after sonication, buffer C was exchanged for FA buffer (50 mM HEPES-KOH, pH 7.5, 140 mM sodium chloride, 1 mM EDTA, 1% v/v Triton X-100, 0.1% w/v sodium deoxycholate, 0.1% w/v SDS, 1X Roche Complete EDTA-free protease inhibitor cocktail, 25 nM MG132 in nuclease-free water) prior to antibody pulldown using Amicon columns (10,000 kDa MWCO). After genomic DNA was isolated for both specific IPs and input DNA, libraries were prepared for paired-end multiplexed tag Solexa/Illumina sequencing using the Illumina DNA Sample Prep Kit v2.0 according to manufacturer instructions.

The following antibodies were used for ChIP-Seq: H3K4me3 (mouse monoclonal, abcam ab1012), H3K27me3 (rabbit polyclonal, Millipore 07-449), H3K9me2 (mouse monoclonal, abcam ab1220). Two replicates were processed, normalized and the data combined for further analyses.

Comparisons with Hi-C data

Processed and normalized MCF-10A Hi-C interaction matrices at 250 kb resolution were from our previous study (Barutcu et al., 2015) and are available in the GEO Database (GSE66733). The overall decay of Hi-C interaction matrices were calculated using a modified LOWESS method ($\alpha = 1\%$, interquartile range filter), as described previously (Sanyal et al., 2012). Then, the Z-scores of the interaction matrices were calculated (Zhang et al., 2012); these are normalized for genomic distance and signal bias. The Pearson correlations were then calculated on the Z score matrices. By performing Principal Component Analysis (PCA), the first principal component usually detects the compartmentalization of the genome. By performing PCA, though, the A-type compartment may result in either a positive or negative eigenvector. To detect which compartment is A-type and which is B-type, the genome-wide gene density was calculated to assign the A-type and B-type compartmentalization. The enrichment of nuclear-matrix-associated-DNA/input ratio was calculated by intersecting the A-type and B-type compartment regions with nuclear-matrix-associated-DNA enrichment values.

Analysis of NGS data

Read mapping—100bp paired-end reads were mapped to the hg19 build of the human genome using Bowtie2 (Langmead and Salzberg, 2012).

Read conversion—Mapped reads were converted to BED-format reads for analysis using SAMTools (Li et al., 2009) and BedTools (Quinlan and Hall, 2010).

GC content—GC content was measured using custom Perl scripts reliant upon BioPerl (Stajich et al., 2002) to calculate GC content from FASTA-format sequence files.

Giemsa chromosome bands—Coordinates of Giemsa bands and RefSeq gene positions were extracted from the UCSC Genome Browser for hg19 (Meyer et al., 2013).

Scatterplots and correlation coefficient calculation—Hexagonal plots and Pearson R correlation coefficient calculations were done in R using the “hexbin” package and the “cor.test” function.

Heat maps and clustering—K-means clustering and heat maps were generated in R. Heat maps were generated using the “heatmap.2” function from the “gplots” library. K-means clustering was performed using the “kmeans” function of the “stats” core package.

Affymetrix arrays and analysis—Human Gene 1.0ST arrays from Affymetrix were used to measure gene expression levels in MCF-10A and MDA-MB-231 cells. cRNA amplification and hybridization to the array was performed by the UMASS Medical School Genomic Core as previously described (Dowdy et al., 2010).

Analysis was performed in R to execute normalization, quality control, transcript-level reporting, annotation, and contrast tests. The “affy” (Gautier et al., 2004) package was used to read in raw fluorescent values from arrays; values were normalized across all arrays using quantile normalization (Bolstad et al., 2003), robust means average (RMA) background correction and median polish (Irizarry et al., 2003). Quality control plots were generated to ensure the arrays did not have any artifacts and post-processing values were in similar ranges. Transcripts were annotated using the “annotate” package, the “hugene10sttranscriptcluster.db” package. Contrast tests were generated and performed using the “limma” package.

Circos plot—Circos plot of NM-DNA-enrichment on human chromosome 11 was done using Circos (Krzywinski et al., 2009).

RESULTS

Crosslinking eliminates artifacts in global MAR identification

Profiling nuclear matrix-attached regions (MARs) across the genome requires a valid method for separating matrix-attached DNA from -unattached fragments. To avoid the technical issues of DNA movement during MAR (or SAR) isolation seen with existing halo-based techniques, we adopted a method in which nuclear structures, chromatin and the nuclear matrix, are cross-linked together before fragmenting chromatin with restriction nucleases and removing DNA fragments unattached to the nuclear matrix (Nickerson et al., 1997). Formaldehyde cross links are then reversed and DNA is evaluated by deep sequencing. We have developed a computational pipeline for high-confidence positional mapping of reads to the genome to precisely map specific nuclear matrix attachment sequences. The protocol is presented in Materials and Methods, and summarized in Figs. 1A,B. Our method offers unprecedented preservation of architectural DNA interactions with

the nuclear matrix as judged by ultrastructure (Figs. 1C,D). It is similar to the formaldehyde crosslinking protocols used for chromatin immunoprecipitation that preserve DNA–protein binding interactions in the context of native chromatin, and are used to map specific sequences with the regulatory proteins that recognize them across an entire genome (Albert et al., 2007; Barski et al., 2007). DNA, as measured by DAPI staining, is effectively removed (Fig. 1C).

MAR-Seq validation using immunofluorescence microscopy

We validated our optimized protocol by immunofluorescence localization of nuclear matrix–associated structures before and after removal of unattached chromatin in MCF-10A and MDA-MB-231 cells (Fig. 1D). Cajal bodies (marked by Coilin), histone locus bodies adjacent to Cajal bodies (marked by NPAT), PML bodies (marked by PML), RNA splicing speckles (marked by SC-35), nucleoli (marked by UBF upstream binding factor), and transcription factories (marked by RNA polymerase II) were preserved in locations that were unaltered by removal of most chromatin from the formaldehyde-crosslinked nucleus.

Evaluating MARs in both normal-like breast and breast cancer cells

Here we present a first demonstration of these methods to profile nuclear matrix–attached DNA across the genome in a comparison of nuclear matrix attachment sites in non-tumorigenic MCF-10A mammary epithelial cells and aggressive, metastatic, MDA-MB-231 breast cancer cells. We identified nuclear matrix–attached DNA in both gene-rich/actively-transcribed regions as well as gene-poor/weakly-transcribed regions. To evaluate the consequences of differences in nuclear matrix–driven organization of the genome, we also profiled gene expression and the positional enrichment of gene-regulatory histone modifications: H3K4me3, H3K27me3, and H3K9me2.

MAR enrichment reciprocally correlates with gene density and expression in breast and breast cancer cells

The staining of metaphase chromosomes with the Giemsa dye (Giemsa, 1904) identifies darker AT-rich heterochromatic regions that are relatively gene-poor (G-bands) and lighter bands (R-bands) that are euchromatic and GC-rich. In human lymphoblasts, nuclear matrix–attached DNA sequences obtained using traditional methods that use high salt (MARs) or low salt (SARs; scaffold attachment regions) extraction protocols are both reportedly enriched in G-bands (Craig et al., 1997). In contrast, DNA sequences retained with a nucleoskeleton after nuclease digestion and isotonic electroelution are enriched in R-bands (Craig et al., 1997).

Nuclear matrix–attached DNA sequences obtained using the method we developed showed a highly significant positive correlation ($r = 0.45$) with gene density as determined by Giemsa staining for MCF-10A cells (Fig. 2A). That is, Giemsa R-bands with a higher density of genes had a higher enrichment in nuclear matrix attachment. We corroborated this result by evaluating gene expression using whole-genome array analysis and found that bands with higher gene expression levels were also significantly enriched in nuclear matrix attachments in MCF-10A cells ($r = 0.32$; Fig. 2B). In contrast, these correlations were reversed for MDA-MB-231 cells, where Giemsa G-bands, with a lower gene density, were more enriched

in nuclear matrix attachments ($r = -0.58$; Fig. 2D) and exhibited lower gene expression levels ($r = -0.32$; Fig. 2E). This was a first indication that nuclear matrix attachment is associated with gene-rich euchromatic regions in MCF-10A cells, but with gene-poor, heterochromatic regions in MDA-MB-231 cells.

Comparing relative GC-content with nuclear matrix enrichment (Fig. 2C,F), we also observed differences between MCF-10A and MDA-MB-231 cells. In MCF-10A cells, GC content was weakly but significantly correlated with nuclear matrix-association ($r = 0.12$). However, in MDA-MB-231 cells, the regions most associated with the nuclear matrix were AT-rich; GC-content and matrix-enrichment were significantly negatively correlated ($r = -0.24$). In the human genome, gene-rich regions are GC-rich or GC-neutral, while gene-poor regions are AT-rich (Lander et al., 2001), so our findings further support a difference between MCF-10A and MDA-MB-231 cells in the relationships between gene expression and nuclear matrix association.

Differential MAR positioning relative to coding regions in normal-like and cancer cells

Following up on these observations, we mapped the positions of nuclear matrix-attached sequences across genes in 200 bp intervals. In MCF-10A cells (Fig. 3A), nuclear matrix-attached sequences for expressed genes were enriched upstream and downstream of the gene body, which starts at the transcription start site and ends at the transcription termination site. There was a highly significant increase in this enrichment for expressed genes compared to non-expressed genes ($p\text{-value} < 10^{-10}$). For non-expressed genes, however, nuclear matrix attachments were more enriched within gene bodies than in flanking sequences.

In MDA-MB-231 cells, nuclear matrix attachments (Fig. 3B) were not enriched in regions flanking the gene body and the differences between expressed and non-expressed genes in these flanking regions were small. Within gene bodies in MDA-MB-231 cells, nuclear matrix attachments were more enriched toward the transcription start site for expressed genes but more enriched toward the transcription termination sites for non-expressed genes.

Chromatin modification enrichment reflects gene expression levels independently of nuclear matrix attachment

Further, we examined the positions of activating or repressive chromatin modifications relative to genes (Fig. 3). We performed ChIP-Seq for histone H3K4me3 (active chromatin), H3K9me2 (inactive), and H3K27me3 (inactive). H3K9 methylation recruits the HP1 proteins (Canzio et al., 2014) while H3K27 methylation recruits Polycomb complexes (Conway et al., 2015), each providing a mechanism for long-term gene silencing. We compared the positional enrichment of these histone modifications relative to gene bodies across the genome to nuclear matrix-attached DNA patterns in both MCF-10A and MDA-MB-231 cells. The positional enrichment of H3K4me3, H3K9me2, and H4K27me3 in relation to gene expression levels was very similar in the two cell lines, indicating that these epigenetic histone modifications exerted the expected influence on gene expression. The activating H3K4me3 modification was more enriched in the gene bodies of actively transcribed genes, with the highest concentrations near the transcription start site, than in non-transcribed genes where the highest levels were near the transcription termination site.

Repressive modifications H3K9me2 and H4K27me3 were, as expected, more enriched in the gene bodies of non-expressed genes in both cell lines. For specific genes, we found both nuclear matrix attachments and activating or repressive histone modifications, making it likely that matrix-attached DNA is nucleosomal (Fig. 4).

To probe the relationships between genomic enrichment of nuclear matrix-attached DNA and of these same histone marks, we took two unbiased approaches, k-means clustering (Fig. 4A) and Pearson *r* correlation (Fig. 4B). The enrichment of matrix-attached DNA, H3K4me3, H3K9me2, and H4K27me3 in 200 bp, non-overlapping genome-wide sequences was measured in both MCF-10A and MDA-MB-231 cells. Combining observations from both approaches, we find that the enrichment of each activating and repressive histone modification is modestly correlated between the cell lines, which is expected based on differences in gene expression patterns. However, we see essentially no correlation (positive or negative) between nuclear matrix-attached DNA enrichment and any of these histone marks. This study is the first in which the positional enrichment of histone modifications has been directly compared to the positional enrichment of matrix-attached DNA on a genome-wide scale.

A Circos histogram (Fig. 5A) summarizes our data for chromosome 11, plotting the differential nuclear matrix-attached sequence enrichments in MCF-10A cells (orange) and MDA-MB-231 cells (blue). MDA-MB-231 cells are known to have a chromosome 11 rearrangement (Satya-Prakash et al., 1981), specifically a chromosome 11-15 translocation (ATCC, 2012) (der(15)t(11;15)(q21;p11.2). This should have no effect on local attachments to the nuclear matrix. Although most enriched regions in MDA-MB-231 cells are also enriched in MCF-10A cells, there is an inverse pattern of the highest levels of enrichment for the two cell lines. These differences often align with Giemsa banding, consistent with our observation of a positive correlation between nuclear matrix attachment and gene density per Giemsa chromosome band for MCF-10A cells (Fig. 2A) and a negative correlation in MDA-MB-231 cells where Giemsa bands with a lower density of genes were more enriched in nuclear matrix attachments (Fig. 2D). A model consistent with our experimental results is presented in Fig. 5B.

Nuclear matrix attached DNA is functionally compartmentalized in open (A-type) chromatin

Recent Hi-C studies have revealed that the genome is partitioned into megabase-scale genomic compartments that are either euchromatic, gene-rich, and highly transcribed (A-type compartments) or heterochromatic, gene-poor, and silent (B-type compartments) (Lieberman-Aiden et al., 2009). The long-range looping interactions within an open or a closed compartment are more frequent than the interactions between open and closed compartments. The open (A-type) compartments preferentially and spatially cluster in the nuclear interior, whereas the closed (B-type) compartments cluster around the nuclear periphery (Cremer and Cremer, 2010). In order to assess whether nuclear matrix-attached DNA domains are associated with the functional compartmentation of the genome, we analyzed a recently published Hi-C dataset for MCF-10A cells (Barutcu et al., 2015). We found a depletion of nuclear matrix attachments in B-type (closed) genomic compartments and an enrichment in A-type (open) compartments (Fig. 5C,D). Indeed, the enrichment of

nuclear matrix bound DNA in A-type compartments was highly statistically significant ($p < 10^{-50}$). These results are consistent with the enrichment of matrix attached sequences in gene bodies (Fig. 3A).

DISCUSSION

Our understanding of how genomic DNA is spatially organized in a eukaryotic cell increased dramatically with the advent of deep sequencing technologies, fluorescence in situ hybridization (FISH) (Lichter et al., 1988; Pinkel et al., 1988), chromosome conformation capture techniques (Barutcu et al., 2015; Lieberman-Aiden et al., 2009), and genome modeling (Imakaev et al., 2015). These technological advances address, in complementary ways, the important question of how the higher-order structure of DNA facilitates the cell-type specific transcription of phenotypic genes, but they had not yet been applied to chromatin organization on the nuclear matrix. In a key step forward, we have developed MAR-Seq, a robust approach for identifying nuclear matrix-attached sequences genome wide. In a first application, we profiled these sequences in a non-tumorigenic mammary epithelial cell and an aggressive metastatic breast cancer cell.

Nuclear matrix attachments are proposed to architecturally organize chromatin, with context-specific activation or silencing of genes. For specific genes, there are examples where the positioning of a gene at the nuclear matrix increases transcription. For example, in “halo” preparations where histones are stripped by high salt extraction and loops of DNA remain attached to the nuclear matrix, active genes are localized to the bases of loops while inactive genes are peripheral (Gerdes et al., 1994). Studies on different specific genes and using different methods for identifying MARs or SARs have not given a clear picture of the relationships between matrix attachment and transcriptional control. A comparative study of high salt and lithium 3,5-diiodosalicylate methods for identifying MARs and SARs, respectively, concluded that SARs are facultative and are found flanking genes that are actively expressed, while MARs are distal to actively-expressed genes and are within transcriptionally silent genes (Linnemann et al., 2009). However, without some form of stabilization of matrix-DNA interactions, these halo-based methods do not preserve the structural integrity of the *in situ* nuclear matrix and may give a very incomplete picture of higher order chromatin organization and its relationship to gene expression. We have addressed these concerns, including the possible sliding of DNA through matrix attachment sites (Mirkovitch et al., 1984) by extensively crosslinking chromatin and nuclear matrix *in situ* (Nickerson et al., 1997) to identify matrix-attached sequences. Our results indicate that the new MAR-Seq method captures both MARs and SARs; gene bodies were found to be enriched in NMADs as were regions directly flanking actively-expressed genes (Fig. 3). Notably MAR-Seq successfully identified NMADs in gene-poor regions of the genome (Fig. 2A,D).

MAR-Seq identified functionally significant differences in genomic regions enriched for nuclear matrix-attachment in normal-like mammary and breast cancer cells. Our results suggest that the nuclear matrix-associated organization of the genome is highly cell-context dependent, consistent with architectural parameters established by chromosome conformation capture (Barutcu et al., 2015). Our method provides new capabilities to define

contributions of the nuclear matrix to compartmentalization of the genome and its regulatory machinery that are obligatory for biological control and compromised in disease.

Accession Numbers

The accession number for the new data reported in this paper is GEO: GSE87671.

Acknowledgements

We thank Lori A. Martin-Buley for editorial assistance and Jennifer Díaz for help with manuscript formatting and preparation.

Contract grant sponsor: National Institutes of Health

Contract grant numbers: P01 CA082834; R37 DE012528

LITERATURE CITED

- Albert I, Mavrich TN, Tomsho LP, Qi J, Zanton SJ, Schuster SC, Pugh BF. Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature*. 2007; 446:572–576. [PubMed: 17392789]
- Ali SA, Dobson JR, Lian JB, Stein JL, van Wijnen AJ, Zaidi SK, Stein GS. A RUNX2-HDAC1 co-repressor complex regulates rRNA gene expression by modulating UBF acetylation. *J Cell Sci*. 2012; 125:2732–2739. [PubMed: 22393235]
- Ali SA, Zaidi SK, Dobson JR, Shakoori AR, Lian JB, Stein JL, van Wijnen AJ, Stein GS. Transcriptional corepressor TLE1 functions with Runx2 in epigenetic repression of ribosomal RNA genes. *Proc Natl Acad Sci USA*. 2010; 107:4165–4169. [PubMed: 20160071]
- ATCC. Thawing, Propagating, and Cryopreserving Protocol NCI-PBCF-HTB26 (MDA-MB-231) Breast Adenocarcinoma. In: Collection ATC. , editor. Physical Sciences-Oncology Center Cetwork Bioresource Core Facility. ATCC; Manassas VA, USA: 2012. Version 1.5 ed
- Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. High-resolution profiling of histone methylations in the human genome. *Cell*. 2007; 129:823–837. [PubMed: 17512414]
- Barutcu AR, Lajoie BR, McCord RP, Tye CE, Hong D, Messier TL, Browne G, van Wijnen AJ, Lian JB, Stein JL, Dekker J, Imbalzano AN, Stein GS. Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. *Genome Biol*. 2015; 16:214. [PubMed: 26415882]
- Benyajati C, Worcel A. Isolation, characterization, and structure of the folded interphase genome of *Drosophila melanogaster*. *Cell*. 1976; 9:393–407. [PubMed: 825231]
- Bode J, Kohwi Y, Dickinson L, Joh T, Klehr D, Mielke C, Kohwi-Shigematsu T. Biological significance of unwinding capability of nuclear matrix-associating DNAs. *Science*. 1992; 255:195–197. [PubMed: 1553545]
- Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003; 19:185–193. [PubMed: 12538238]
- Boulikas T. Chromatin domains and prediction of MAR sequences. *Int Rev Cytol*. 1995; 162A:279–388. [PubMed: 8575883]
- Canzio D, Larson A, Narlikar GJ. Mechanisms of functional promiscuity by HP1 proteins. *Trends Cell Biol*. 2014; 24:377–386. [PubMed: 24618358]
- Conway E, Healy E, Bracken AP. PRC2 mediated H3K27 methylations in cellular identity and cancer. *Curr Opin Cell Biol*. 2015; 37:42–48. [PubMed: 26497635]
- Cook PR, Brazell IA. Supercoils in human DNA. *J Cell Sci*. 1975; 19:261–279. [PubMed: 1202042]
- Cook PR, Brazell IA, Jost E. Characterization of nuclear structures containing superhelical DNA. *J Cell Sci*. 1976; 22:303–324. [PubMed: 1002771]

- Craig JM, Boyle S, Perry P, Bickmore WA. Scaffold attachments within the human genome. *J Cell Sci.* 1997; 110:2673–2682. Pt 21. [PubMed: 9427385]
- Cremer T, Cremer M. Chromosome territories. *Cold Spring Harb Perspect Biol.* 2010; 2:a003889. [PubMed: 20300217]
- Davie JR. The nuclear matrix and the regulation of chromatin organization and function. *Int Rev Cytol.* 1995; 162A:191–250. [PubMed: 8575881]
- Debnath J, Mills KR, Collins NL, Reginato MJ, Muthuswamy SK, Brugge JS. The role of apoptosis in creating and maintaining luminal space within normal and oncogene-expressing mammary acini. *Cell.* 2002; 111:29–40. [PubMed: 12372298]
- Dekker J. Two ways to fold the genome during the cell cycle: insights obtained with chromosome conformation capture. *Epigenetics Chromatin.* 2014; 7:25. [PubMed: 25435919]
- Dowdy CR, Xie R, Frederick D, Hussain S, Zaidi SK, Vradii D, Javed A, Li X, Jones SN, Lian JB, van Wijnen AJ, Stein JL, Stein GS. Definitive hematopoiesis requires Runx1 C-terminal-mediated subnuclear targeting and transactivation. *Hum Mol Genet.* 2010; 19:1048–1057. [PubMed: 20035012]
- Fey EG, Krochmalnic G, Penman S. The nonchromatin substructures of the nucleus: the ribonucleoprotein (RNP)-containing and RNP-depleted matrices analyzed by sequential fractionation and resinless section electron microscopy. *J Cell Biol.* 1986; 102:1654–1665. [PubMed: 3700470]
- Flickinger RA. Localization of AT-rich sequences at the nuclear matrix. *Cell Biol Int Rep.* 1986; 10:415–420. [PubMed: 3527451]
- Gautier L, Cope L, Bolstad BM, Irizarry RA. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics.* 2004; 20:307–315. [PubMed: 14960456]
- Gerdes MG, Carter KC, Moen PT Jr, Lawrence JB. Dynamic changes in the higher-level chromatin organization of specific sequences revealed by in situ hybridization to nuclear halos. *J Cell Biol.* 1994; 126:289–304. [PubMed: 8034736]
- Ghule PN, Dominski Z, Lian JB, Stein JL, van Wijnen AJ, Stein GS. The subnuclear organization of histone gene regulatory proteins and 3' end processing factors of normal somatic and embryonic stem cells is compromised in selected human cancer cell types. *J Cell Physiol.* 2009; 220:129–135. [PubMed: 19277982]
- Gibcus JH, Dekker J. The hierarchy of the 3D genome. *Mol Cell.* 2013; 49:773–782. [PubMed: 23473598]
- Giemsa G. Eine vereinfachung und vervollkommnung meiner methylenazur-methylenblau-eosin-färbemethode zur erzielung der Romanowsky-Nocht'schen chromatinfärbung. *Zentabl Bakteriell Parasitenkd Infectkrankh.* 1904; 37:308.
- He DC, Nickerson JA, Penman S. Core filaments of the nuclear matrix. *J Cell Biol.* 1990; 110:569–580. [PubMed: 2307700]
- Imakaev MV, Fudenberg G, Mirny LA. Modeling chromosomes: Beyond pretty pictures. *FEBS Lett.* 2015; 589:3031–3036. [PubMed: 26364723]
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics.* 2003; 4:249–264. [PubMed: 12925520]
- Jackson DA, Dickinson P, Cook PR. The size of chromatin loops in HeLa cells. *EMBO J.* 1990; 9:567–571. [PubMed: 2303042]
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009; 19:1639–1645. [PubMed: 19541911]
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann Y, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L,

Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissole SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglu S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kasprzyk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler J, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrino A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, Szustakowski J, International Human Genome Sequencing C. Initial sequencing and analysis of the human genome. *Nature*. 2001; 409:860–921. [PubMed: 11237011]

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9:357–359. [PubMed: 22388286]

Lee TI, Johnstone SE, Young RA. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat Protoc*. 2006; 1:729–748. [PubMed: 17406303]

Leonhardt H, Rahn HP, Weinzierl P, Sporbert A, Cremer T, Zink D, Cardoso MC. Dynamics of DNA replication factories in living cells. *J Cell Biol*. 2000; 149:271–280. [PubMed: 10769021]

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25:2078–2079. [PubMed: 19505943]

Lichter P, Cremer T, Borden J, Manuelidis L, Ward DC. Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Hum Genet*. 1988; 80:224–234. [PubMed: 3192212]

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009; 326:289–293. [PubMed: 19815776]

Linnemann AK, Platts AE, Krawetz SA. Differential nuclear scaffold/matrix attachment marks expressed genes. *Hum Mol Genet*. 2009; 18:645–654. [PubMed: 19017725]

Ma T, Van Tine BA, Wei Y, Garrett MD, Nelson D, Adams PD, Wang J, Qin J, Chow LT, Harper JW. Cell cycle-regulated phosphorylation of p220(NPAT) by cyclin E/Cdk2 in Cajal bodies promotes histone gene transcription. *Genes Dev*. 2000; 14:2298–2313. [PubMed: 10995387]

Maya-Mendoza A, Aranda-Anzaldo A. Positional mapping of specific DNA sequences relative to the nuclear substructure by direct polymerase chain reaction on nuclear matrix-bound templates. *Anal Biochem*. 2003; 313:196–207. [PubMed: 12605856]

Maya-Mendoza A, Hernandez-Munoz R, Gariglio P, Aranda-Anzaldo A. Gene positional changes relative to the nuclear substructure correlate with the proliferating status of hepatocytes during liver regeneration. *Nucleic Acids Res*. 2003; 31:6168–6179. [PubMed: 14576303]

- Meyer LR, Zweig AS, Hinrichs AS, Karolchik D, Kuhn RM, Wong M, Sloan CA, Rosenbloom KR, Roe G, Rhead B, Raney BJ, Pohl A, Malladi VS, Li CH, Lee BT, Learned K, Kirkup V, Hsu F, Heitner S, Harte RA, Haeussler M, Guruvadoo L, Goldman M, Giardine BM, Fujita PA, Dreszer TR, Diekhans M, Cline MS, Clawson H, Barber GP, Haussler D, Kent WJ. The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res.* 2013; 41:D64–69. [PubMed: 23155063]
- Mirkovitch J, Mirault ME, Laemmli UK. Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. *Cell.* 1984; 39:223–232. [PubMed: 6091913]
- Mirkovitch J, Spierer P, Laemmli UK. Genes and loops in 320,000 base-pairs of the *Drosophila melanogaster* chromosome. *J Mol Biol.* 1986; 190:255–258. [PubMed: 3098982]
- Nelson WG, Pienta KJ, Barrack ER, Coffey DS. The role of the nuclear matrix in the organization and function of DNA. *Annu Rev Biophys Chem.* 1986; 15:457–475. [PubMed: 3013231]
- Nickerson JA, Krockmalnic G, Wan KM, Penman S. The nuclear matrix revealed by eluting chromatin from a cross-linked nucleus. *Proc Natl Acad Sci USA.* 1997; 94:4446–4450. [PubMed: 9114009]
- Pathak RU, Srinivasan A, Mishra RK. Genome-wide mapping of matrix attachment regions in *Drosophila melanogaster*. *BMC Genomics.* 2014; 15:1022. [PubMed: 25424749]
- Paulson JR, Laemmli UK. The structure of histone-depleted metaphase chromosomes. *Cell.* 1977; 12:817–828. [PubMed: 922894]
- Pinkel D, Landegent J, Collins C, Fuscoe J, Segraves R, Lucas J, Gray J. Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. *Proc Natl Acad Sci USA.* 1988; 85:9138–9142. [PubMed: 2973607]
- Pombo A, Dillon N. Three-dimensional genome architecture: players and mechanisms. *Nat Rev Mol Cell Biol.* 2015; 16:245–257. [PubMed: 25757416]
- Pratap J, Wixted JJ, Gaur T, Zaidi SK, Dobson J, Gokul KD, Hussain S, van Wijnen AJ, Stein JL, Stein GS, Lian JB. Runx2 transcriptional activation of Indian Hedgehog and a downstream bone metastatic pathway in breast cancer cells. *Cancer Res.* 2008; 68:7795–7802. [PubMed: 18829534]
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010; 26:841–842. [PubMed: 20110278]
- Razin SV, Gavrilov AA. Chromatin without the 30-nm fiber: constrained disorder instead of hierarchical folding. *Epigenetics.* 2014; 9:653–657. [PubMed: 24561903]
- Razin SV, Gromova II, Iarovaia OV. Specificity and functional significance of DNA interaction with the nuclear matrix: new approaches to clarify the old questions. *Int Rev Cytol.* 1995; 162B:405–448. [PubMed: 8557493]
- Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature.* 2012; 489:109–113. [PubMed: 22955621]
- Satya-Prakash KL, Pathak S, Hsu TC, Olive M, Cailleau R. Cytogenetic analysis on eight human breast tumor cell lines: high frequencies of 1q, 11q and HeLa-like marker chromosomes. *Cancer Genet Cytogenet.* 1981; 3:61–73. [PubMed: 7272986]
- Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, Fuellen G, Gilbert JG, Korf I, Lapp H, Lehvaslaiho H, Matsalla C, Mungall CJ, Osborne BI, Pocock MR, Schattner P, Senger M, Stein LD, Stupka E, Wilkinson MD, Birney E. The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.* 2002; 12:1611–1618. [PubMed: 12368254]
- Vogelstein B, Pardoll DM, Coffey DS. Supercoiled loops and eucaryotic DNA replication. *Cell.* 1980; 22:79–85. [PubMed: 7428042]
- Zhang Y, McCord RP, Ho YJ, Lajoie BR, Hildebrand DG, Simon AC, Becker MS, Alt FW, Dekker J. Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell.* 2012; 148:908–921. [PubMed: 22341456]
- Zhao J, Kennedy BK, Lawrence BD, Barbie DA, Matera AG, Fletcher JA, Harlow E. NPAT links cyclin E-Cdk2 to the regulation of replication-dependent histone gene transcription. *Genes Dev.* 2000; 14:2283–2297. [PubMed: 10995386]

Fig 1.

Isolation of nuclear matrix–attached DNA: workflow and validation. (A) The workflow of nuclear matrix associated DNA and library preparation. (B) Scheme of nuclear matrix–DNA interactions in whole cell/CSK extracted nuclei (left), digested nuclei (right), and after salt extraction (lower). Right - dotted lines indicate DNA cut by restriction enzymes. (C) DIC images and DAPI staining of MDA-MB-231 cells at three stages of isolation of nuclear matrix: whole cell (top), CSK-extracted nuclei digested with restriction enzymes (middle), and nuclear matrix (bottom). Note that DAPI exposure is equal across images. (D) Confocal

immunofluorescence micrographs of critical nuclear structures in MCF-10A and MDA-MB-231 cells after CSK extraction (CSK) or following full nuclear matrix isolation (NM). Images are representative of results seen in 5 independently replicated experiments. Markers of nuclear matrix-associated structures include: Coilin (Cajal bodies), NPAT (nuclear protein ataxia-telangiectasia locus - histone locus bodies), PML (promyelocytic leukemia - PML bodies), SC-35 (serine/arginine-rich splicing factor 2 - RNA splicing speckles), UBF (upstream binding transcription factor, RNA polymerase I - nucleoli), and Pol II (RNA polymerase II - mRNA synthesis sites). Shown are representative images from a total of five replicate experiments. Note that nuclear structure and subnuclear domains are well preserved through the isolation of the crosslink-stabilized nuclear matrix.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 2.

Cytogenetic-scale relationships of nuclear matrix–attached DNA with genetic architecture and gene expression differ between MCF-10A and MDA-MB-231 cells. (A,D) Scatterplots of percent of gene coverage per Giemsa cytoBand/chromosome band (y-axis) and nuclear matrix–attached DNA (NMAD) enrichment per cytoBand (x-axis) in MCF-10A cells (A) and MDA-MB-231 cells (D). These x-y pairs are plotted for the entire genome using a hexagonal density scatterplot in which an increased number of shared x-y observations are represented as darker hexagons. (B,E) In MCF-10A (B) and MDA-MB-231 cells (E) the following was done: for each cytoBand, the mean gene level normalization (Robust Multi-array Average (RMA)-normalized transcript detection value (Irizarry et al., 2003) extracted from Affymetrix Human Gene 1.0ST arrays) for all genes found within the given cytoBand was plotted as the y-coordinate and the corresponding enrichment of NMAD was plotted as the x-coordinate. Array analysis data are combined, normalized results from two biological replicates. (C,F) Sequences associated with the nuclear matrix were modestly GC-enriched in MCF-10A cells but AT-enriched in MDA-MB-231 cells. Using genome-wide non-overlapping 200 bp intervals, nuclear matrix-associated DNA (NMAD) enrichment was measured in MCF-10a cells (C) and MDA-MB-231 cells (F) and is represented on the x-axis. For each 200 bp interval, the corresponding percent of G plus C bases within the interval is plotted as the dependent variable on the y-axis. These x-y pairs are plotted for the entire genome using a hexagonal density scatterplot where an increased number of shared x-y observations are represented as darker hexagons. (A-F) The Pearson r- and p-values were calculated for all analyses shown here and the line of best fit was plotted in red over the plots; the Pearson r values are indicated. P-values in all cases were calculated as $<2.2 \times 10^{-16}$.

Fig. 3.

Analysis of histone marks and nuclear matrix-attached DNA. Two biological replat samples of MCF-10A (A) and MDA-MB-231 cells (B) were analyzed for nuclear matrix attachment sites and the indicated epigenetic histone modifications. All deep sequencing data analyzed here are combined, normalized results from the replicates. For each annotated transcript in RefSeq, the enrichment (IP/Input) of H3K4me3, H3K27me3, H3K9me2, and nuclear matrix-attached DNA (NMAD) as indicated over each panel, was measured in 30 defined relative genomic regions: 10 kb 5'-transcriptional start site (TSS) to the TSS in 1 kb steps, transcription termination site (TTS) to 10 kb 3'-TTS in 1Kb steps, and TSS to TTS in 10 approximately equal intervals (scaled to 1 kb in length). Using a Wilcoxon rank sum test, enrichment proximal to genes expressed in a cell line was contrasted with enrichment proximal to genes not expressed in a cell line. Lines represent the mean enrichment of the epigenetic mark being measured (left y-axis) in the relative genomic region (x-coordinate) for the set of genes being contrasted (color-coded and defined below graphs). Triangles correspond to the approximate p-value (95% CI) for the Wilcoxon rank sum test run at each relative interval (right y-axis).

Fig. 4. On a genome-wide scale, nuclear matrix-associated DNA (NMAD) does not cluster with histone modifications that regulate gene expression. (A) Using genome-wide, non-overlapping 200 bp intervals, the enrichment of NMAD (NMAD/Digested Nuclei), and the enrichment of the histone modifications H3K4me3, H3K27me3, and H3K9me2 (IP/Input) were measured and plotted as a heat map. Each horizontal row is the same 200 bp interval. For each nuclear feature measured (columns), the $\log_2$ of the enrichment values was plotted and then clustered using the k-means expectation-maximization algorithm. The color scale

for log2 (enrichment) is located below the heat map. The colored bar on the left indicates each cluster. (B) Pairwise matrix of scatterplots comparing the enrichment of each nuclear mark in each cell line. Scatterplots (upper left panels) are hexagonal density scatter plots where darker dots indicate an increased number of paired x-y observations. Pearson r values were calculated for each pairwise comparison. The line of best fit is plotted in red on the scatterplots.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 5.

Patterns of nuclear matrix-attached DNA enrichment in MCF-10A and MDA-MB-231 cells show a differential association with transcription. (A) As shown in Figs. 2 and 3, nuclear matrix attachment is not associated with transcription or active histone marks in MDA-MB-231 cells. This difference is summarized for Chromosome 11 in the Circos histogram plot of nuclear matrix enrichment for MDA-MB-231 cells (blue) compared to MCF-10A cells (orange). The outer ring depicts ranges of Giemsa stain positivity per chromosomal band; red = centromere, white to black (and intermediate grays) indicate intensity of Giemsa staining (Giemsa, 1904). Innermost ring of black bars indicates positions of gene bodies. (B) Model depicting aberrant DNA-matrix interactions in MDA-MB-231 cells leading to alterations in gene expression. (C) In MCF-10A cells, nuclear matrix-attached DNA is enriched at open (A-type) genomic compartments. Shown is a UCSC genome browser snapshot of chromosome 1 showing the genomic compartmentalization (1st eigen vector) of the MCF-10A cell line assessed by Hi-C (Barutcu et al., 2015) (top), enrichment of nuclear matrix associated DNA (NMAD) over the input control (middle panel), and the UCSC genes (lower panel). (D) Box plot for MCF-10A showing the genome-wide quantification of the nuclear matrix-attached DNA enrichment in A-type (open) compartment regions (red), and B-type (closed) compartment regions (blue). The A-type (open) compartment regions are associated with higher enrichment of nuclear matrix-attached DNA, p-value: Wilcoxon rank-sum test. Note that deep sequencing and Hi-C data used for these analysis are combined, normalized results from two biological replicates each.