Differential nuclear scaffold/matrix attachment marks expressed genes[†]

Amelia K. Linnemann[1], Adrian E. Platts[1,2] and Stephen A. Krawetz[1,2,3,*]

[1]The Center for Molecular Medicine and Genetics, [2]Department of Obstetrics and Gynecology and [3]Institute for Scientific Computing, Wayne State University School of Medicine, C.S. Mott Center, 275 E Hancock, Detroit, MI 48201, USA

It is well established that nuclear architecture plays a key role in poising regions of the genome for transcription. This may be achieved using scaffold/matrix attachment regions (S/MARs) that establish loop domains. However, the relationship between changes in the physical structure of the genome as mediated by attachment to the nuclear scaffold/matrix and gene expression is not clearly understood. To define the role of S/MARs in organizing our genome and to resolve the often contradictory loci-specific studies, we have surveyed the S/MARs in HeLa S3 cells on human chromosomes 14–18 by array comparative genomic hybridization. Comparison of LIS (lithium 3,5-diiodosalicylate) extraction to identify SARs and 2 M NaCl extraction to identify MARs revealed that approximately one-half of the sites were in common. The results presented in this study suggest that SARs 5′ of a gene are associated with transcript presence whereas MARs contained within a gene are associated with silenced genes. The varied functions of the S/MARs as revealed by the different extraction methods highlights their unique functional contribution.

INTRODUCTION

Within the nucleus chromosomes occupy distinct territories (1) from which actively transcribing genes may extend and loop into structurally distinct interchromatin compartments (2–4). Considerable evidence has accumulated to suggest that the topological constraints required for looping are provided through the association of discrete regions of the genome with the nuclear scaffold/matrix (5) at scaffold/matrix attachment regions, or S/MARs. The nuclear scaffold/matrix provides an anchor for higher order genome structure that is more than a simple mechanical organizer. *Trans*-factors including topoisomerases (6) are often found in association with the nuclear scaffold/matrix. Working in conjunction with other chromatin modifiers they may actively promote chromatin restructuring to reduce torsional stress and activate processes, or conversely, condense and silence various chromosomal segments (7). In this manner, the nuclear scaffold/matrix is dynamically modified during the cell cycle to serve a continuously changing role.

Regions of the genome attach to the nuclear scaffold/matrix in both a cell type and cell cycle context specific manner (8,9). Although the precise mechanism(s) await determination, S/MARs exhibit varied functions that include augmenting transcription (5,10), insulating genic domains (11–13) and facilitating replication (14,15). For example, the positions of MARs of the human β-globin locus are arranged to specifically facilitate developmentally ordered transcription/repression (16). Induction of gene expression at the mouse T-helper 2 cytokine locus is correlated with a local increase in the total number of MARs as they form a series of small loops (17). These and other single-gene locus association studies hint at the importance of the nuclear scaffold/matrix in domain remodeling to permit transcription. However, several loci focused studies of S/MARs on chromosome 16 (6,10,18,19), as well as extended mammalian studies of megabase size genomic regions (20) have yet to reach a consensus.

Transcriptionally active regions as well as regions undergoing replication (21) are segregated into 50–200 kb looped domains through their dynamic association with the nuclear

*To whom correspondence should be addressed. Tel: +1 3135776770; Fax: +1 3135778554; Email: steve@compbio.med.wayne.edu

scaffold/matrix (22–25). Since attachment of the genome to the nuclear scaffold/matrix is dynamic and contextually dependent, the exact number and locations of genomic attachment sites remains contentious. It has been estimated that ~64 000 S/MARs divide the somatic genome into a series of ~100 kb domains, given that each domain is bounded by an S/MAR at each end. However, only a subset of potential S/MARs may be active in a cell at any given time. Changes in the activity of these sites may provide a means to modulate phenotype (8). To date, few S/MARs have been identified and this has hampered efforts to develop meaningful biological models. Attempts to identify S/MARs *in silico* are largely refractory to sequence analysis often yielding over predictive models (reviewed in 26).

Several protocols have been engineered to isolate DNA tethered to the nuclear scaffold/matrix away from freely extended loop DNA. The most widely used methods rely on LIS (lithium 3,5-diiodosalicylate) or 2 M NaCl to isolate what are viewed as different types of attachment sites. LIS appears to disrupt binding mediated through transcription complexes (27), yielding the nuclear scaffold, whereas 2 M NaCl has been suggested to isolate a nuclear matrix interwoven with newly synthesized RNA (28). Accordingly, distinct groups of SARs or MARs should be identified by each method.

The literature contains many small-scale S/MAR studies using different isolation methods as well as a variety of cell types. These have utilized both *in vivo* analysis and *in vitro* reassociation to study scaffold/matrix association potential. Although *in vitro* studies have shown that LIS and 2 M NaCl isolate similar attachment sites congruent with structural analyses (29), the differences in attachment *in vivo* suggest that the nuclear environment plays a larger role than binding potential alone. Indeed, small-scale studies comparing NaCl and LIS extraction-based protocols have been shown to isolate different regions of attachment (30).

To establish the genomic differences between isolation methods and their potential role in gene expression, we have mapped, at the chromosomal level, HeLa S3 MARs and SARs isolated by either 2 M NaCl or LIS extraction respectively using array comparative genomic hybridization (aCGH). The results for chromosome 16 are highlighted as they generalize to chromosomes 14, 15, 17 and 18 (see Supplementary Materials). Nuclear matrices prepared with 2 M NaCl were primarily associated with intergenic gene-poor regions and genes that were attached to the nuclear matrix were silent. Conversely, LIS-isolated SARs were more closely associated with genes with many overlapping the genes themselves. Interestingly, SARs residing within the 5′ proximal region of genes were coupled with higher transcript levels. This first chromosomal-wide survey suggests that SARs and MARs work in concert to mediate genome organization and facilitate expression.

RESULTS

S/MARs were isolated by either 2 M NaCl or 25 mM LIS to remove histones thereby enabling unconstrained DNA to diffuse away from the nuclear scaffold/matrix forming a peripheral halo (Fig. 1A). The halo of unconstrained loop DNA was released from the scaffold/matrix-bound DNA by *Eco*RI restriction enzyme digestion then separated by sedimentation. The total amount of DNA recovered after restriction digestion in both the loop and scaffold/matrix fractions was similar. Subsequent DNase I digestion can reduce the scaffold/matrix fraction to 10% of the estimated total genomic DNA (data not shown). The differential hybridization of the loop and scaffold/matrix fractions were compared using the Nimblegen Systems human whole-genome CGH array system (CGAR0150-WHG8CGH array 7) yielding chromosome-wide profiles of nuclear scaffold and nuclear matrix association.

MAR-mediated chromosomal looping correlates with gene-dense regions

Normalized signal ratios of the loop to scaffold/matrix signal from all aCGH probes for each extraction method from duplicate independent experiments were calculated. Replicates maintained high correlation coefficients (HeLa LIS: $r = 0.859$ and HeLa NaCl: $r = 0.610$) compared with randomly permuted signal ratios (LIS: median: $r = 0.020$, SD = 0.009; NaCl: median: $r = 0.024$, SD = 0.008). To moderate the impact of residual variance, probe signals for the LIS and NaCl extractions were averaged between replicates then assessed as a function of their position along the chromosome (Fig. 1B). LIS scaffold and NaCl matrix association are indicated by negative signal ratios while positive values indicate loop enrichment. The chromosomal organization with respect to the nuclear scaffold/matrix was compared with G-banding and gene density. At the chromosomal level, the positive correlation of gene density with loop enrichment relative to the NaCl prepared nuclear matrix is clear and accentuated at the telomeric regions. In contrast, the AT rich, gene poor, G banded regions are matrix enriched. Analysis of the LIS nuclear scaffold showed no significant relationship with gene density.

Identification and global characterization of S/MARss

Distinct sites of scaffold/matrix attachment were identified using a three tier process to minimize type 1 false-positive error as verified by permutation analysis. The signal was not normally distributed; hence a non-parametric statistic that segregated data analogous to 2 SD above the mean was adopted. Accordingly, only probes exhibiting a signal ratio ($\log_2[\text{loop}/\text{matrix}]$ and $\log_2[\text{loop}/\text{scaffold}]$) in the extreme 2.5% of all signals were considered. This level of signal is analogous to an ~70% enrichment or higher in the scaffold/matrix fraction that should resolve a robust and stable set of attachments. Dynamic or transient attachments were also expected to occur, but only in a subset of cells. These regions would likely resolve within the approximately equal portions of loop and scaffold/matrix, represented by signal ratios near zero. They would be difficult to resolve within this background and as expected they show significant variability with respect to PCR validation and were not considered further. To filter spurious outlier signals, additional pairs of probes with a similar signal distribution within 3 kb either
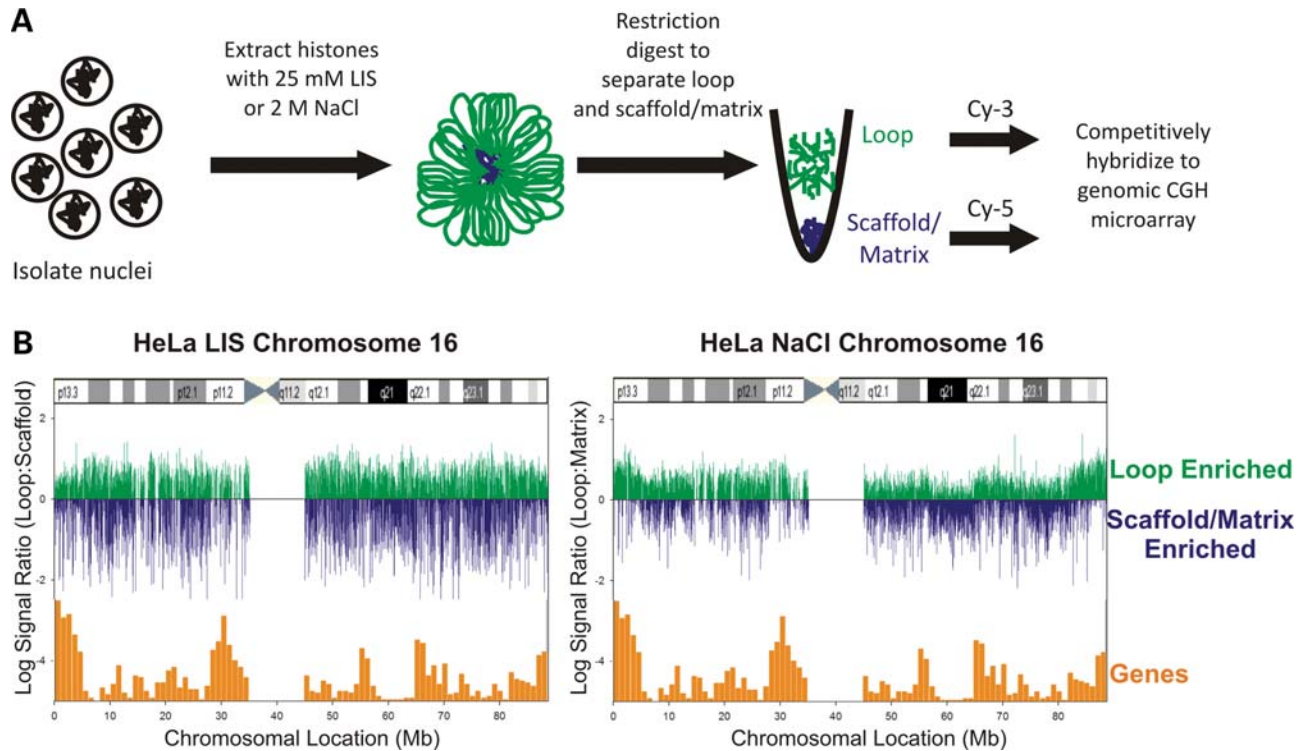
Figure 1. Nuclear Scaffold/Matrix extraction reveals isolation specific differences. (**A**) Isolation of loop and scaffold/matrix DNA. (**B**) HeLa $\log_2$ CGH signal ratios of LIS isolated loop:scaffold and NaCl-isolated loop:matrix fractions. The $\log_2$ CGH ratios are indicated as bars along chromosome 16. Loop regions (green) are indicated as positive signals, whereas nuclear scaffold/matrix-associated regions (blue) are represented as negative signals respective of the zero axis. Gene density (orange bars) and an ideogram representative of G-banding were then overlaid along the chromosome. The NaCl-extracted nuclear matrix revealed global chromatin organization such that gene density correlated with looping. It is apparent that the LIS-isolated nuclear scaffold organizes the chromosome in a different manner.

side of the extreme probe were required. Finally, the average signal across each entire restriction fragment was required to be concordant with the primary observation. Upon meeting these criteria the independent biological replicates were compared, resolving 1016 SARs and 775 MARs along chromosome 16. A total of 403 regions of attachment were shared between both extraction protocols (Supplementary Material, Tables S1 and S2 for complete list along with other chromosomes assayed). The veracity of this strategy for identifying regions of attachment from each extraction method was then validated by quantitative real-time PCR as described (31,32). This revealed a high level of concordance between the scaffold/matrix enrichment assessed on the arrays and that measured by PCR (Supplementary Material, Table S3).

The relative distance between each S/MAR can be used to estimate loop size. The uneven spacing of S/MARs encompassing a range of loop sizes across chromosome 16 is illustrated in Figure 2 (see Supplementary Material, Fig. S1 for the other chromosomes examined). Both SARs and MARs show a bimodal distribution consistent with two ranges of loop sizes that may serve different functions as predicted by the interchromosomal network model that also proposes various classes of nuclear attachment (33). The first peak contained S/MARs that were spaced from 87 to 2217 bp apart, yielding a 762 bp average MAR loop and 558 bp average SAR loop. This represented ~17% of all S/MAR bounded loop domains. S/MARs creating these small loops were distributed

across the entire length of the chromosome with no preference for either gene-dense or gene-poor regions. The larger peak contained ~78% of the loop domains created by S/MARs that were spaced from ~3.3 kb to ~970 kb apart. This subset yielded an average MAR loop of ~94 kb and SAR loop of ~88 kb.

S/MAR sequence properties

Regions of nuclear scaffold/matrix attachment were examined using RegionMiner (Genomatix Software GmbH, Munich, Germany) to determine general S/MAR sequence characteristics. This analysis included the mapping of S/MARs relative to genes and a host of *trans*-factor binding sites as well as assessing conservation within the regions identified. Significance of enrichment within the S/MARs relative to the loop-enriched regions (compare Supplementary Material, Tables S1 and S2) was calculated by chi-square at 95% CI (1 *df*).

MARs tend to be located in intergenic regions (439 of 775 MARs are intergenic) while SARs tend to overlap genes (560 of 1016 MARs are genic). When compared with the loop-enriched regions, the genic/intergenic distributions are statistically significant for both MARs and SARs ($P \leq 0.001$). The majority of MARs and SARs contain at least one conserved region of at least 50 nucleotides (576 of 775 MARs and 767 of 1016 SARs). Their frequency is not significantly different from that observed in the loop-enriched regions. Approxi-
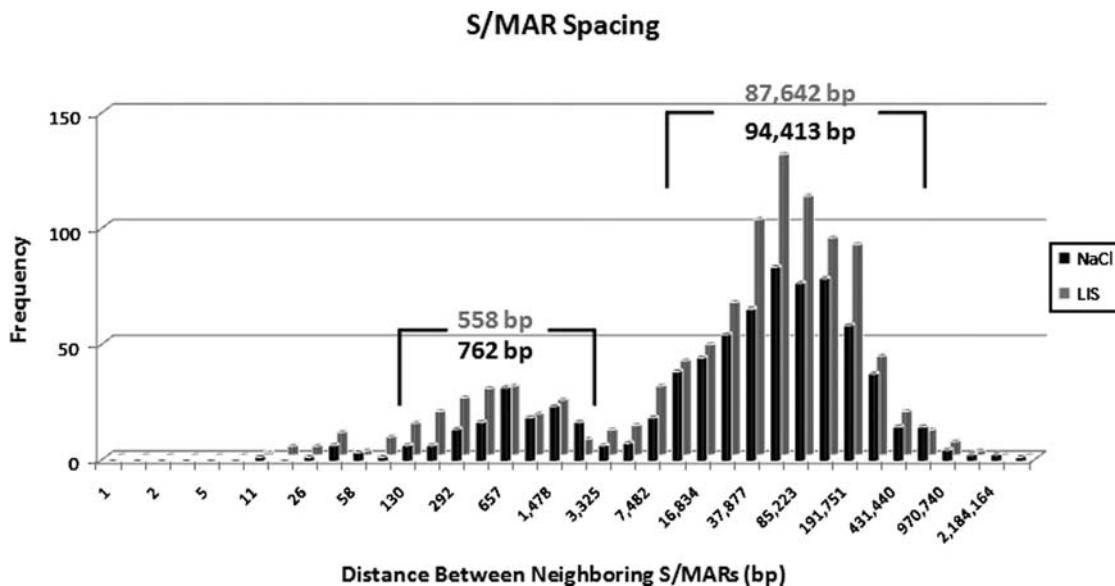
Figure 2. Spacing of S/MARs along human chromosome 16. The binned distance between S/MARs on human chromosome 16 is shown for nuclear scaffolds/matrices isolated by LIS or NaCl extraction with the peak distance averages for each indicated. The frequency of both SAR and MAR spacing shows a bimodal distribution with two groups of average inferred loop sizes at 558 bp and ~88 kb for SARs and 762 bp and ~94 kb for MARs.

mately 51% of the regions that are common between the MAR and SAR data sets overlap genes (207 of 403 regions are genic). All other parameters queried including the AT distribution within the S/MAR fractions were unremarkable.

The distribution of the intergenic S/MARs relative to the nearest gene is summarized in Figure 3 (see Supplementary Material, Fig. S2 for other chromosomes examined). Both intergenic MARs and SARs are similarly distributed from the 5′ and 3′ ends of all genes reflective of the chromosomal distribution of signal observed in Figure 1. Only a subset of the intergenic S/MARs is located immediately proximal to genes. The median distance to the nearest 5′ end MAR was ~207 kb, whereas the 3′ end MAR was located ~126 kb away. In comparison, the median distance to the nearest 5′ end SAR was ~169 kb, whereas the 3′ end SAR was located ~113 kb away. These similarities were reiterated between extraction methods on the various chromosomes analyzed (Supplementary Material, Fig. S2).

Candidate *trans*-factor S/MAR associations

Lamin proteins are a part of the nuclear matrix (34) and MARs specifically bind to matrix components of the nuclear lamina such as lamin B1 (35). Comparison of S/MARs identified in this study with lamin B1-associated domains, LADs (36), revealed significant but varied overlap. For example, ~64% of the MARs and ~51% of the SARs overlap LADs on chromosome 16. However, the level of S/MAR/LAD overlap varies throughout the genome, i.e. chr14: 52/44%; chr15: 41/34%; chr17: 41/26% and chr18: 52/42%. Throughout these regions of overlap, there appear to be clusters of S/MARs within LADs.

The distribution of several known nuclear scaffold/matrix trans-factors or associated families within both the MAR and SAR regions was assessed *in silico* and compared with that

observed in the loop fraction. These included sites for AT-binding factor, CTCF, p53, SWI/SNF-related nucleophosphoproteins, SATB, SOX/SRY-sex/testis determining and related HMG box factors, GC box factors SP1/GC, STAT, Y-box binding transcription factors and YY1. Interestingly, these binding sites were not enriched within the S/MAR fraction when compared with the loop fraction as might have been expected. Of note, both SARs and MARs have a significantly reduced number of CTCF binding sites when compared with loop-enriched regions ($P \leq 0.001$), suggesting that CTCF is not a key player in S/MAR function within HeLa cells as reported in other cases (12).

S/MAR-mediated organization and gene expression

Small-scale studies have suggested that NaCl extracted nuclear matrix preparations will identify MARs that are associated with transcriptionally active regions (28). Similarly, since LIS disrupts transcription complexes (27), the SARs thus isolated were expected to support an indirect function, such as potentiation, to spatially poise rather than direct the interaction of the nuclear scaffold with the transcription factory (37). To test these relationships, the analysis of nuclear scaffold/matrix profiles relative to expressed and silent genes was undertaken. Transcript profiles were established using the Illumina WG8 v2.0 RefSeq bead array system. This interrogated 2885 genes that were assayed by aCGH on chromosomes 14–18, of which 364 were found to be present at a level above the lowest signal value for the Illumina spike-in controls ($S_{min} > 3000$).

As illustrated in Figure 4, gene-dense regions contain both expressed, e.g. MAPK and XTP3TPA, and silent genes, whereas gene-poor regions contain relatively few expressed genes and many silenced genes such as CDH8. The blue chromosomal-wide nuclear scaffold/matrix and green loop
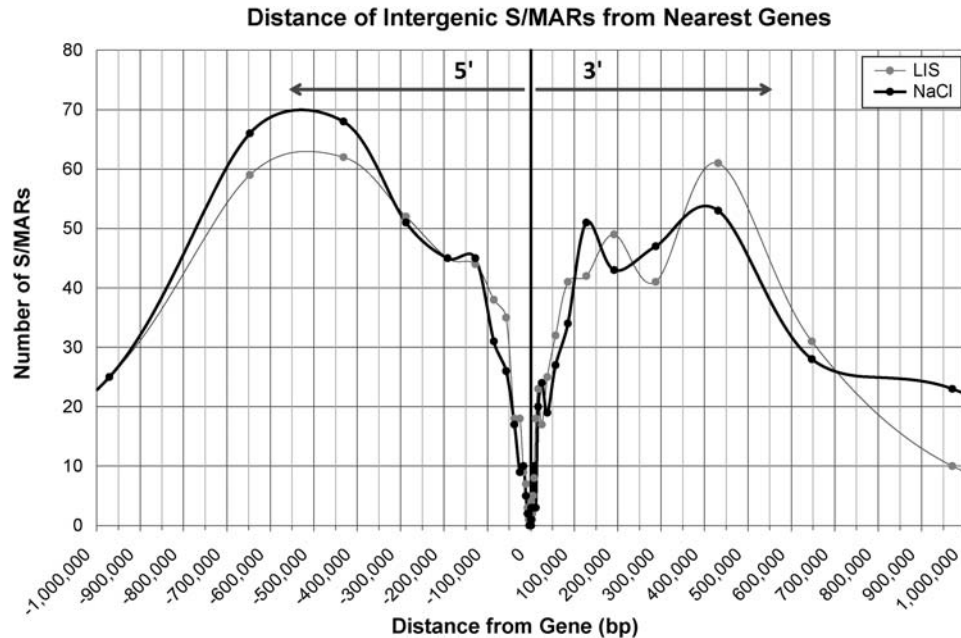
Figure 3. Intergenic nuclear scaffold/matrix attachment and expression. SARs and MARs were mapped according to their distance from both the 5′ and 3′ ends of genes on chromosome 16. Analysis of intergenic S/MAR distance from all genes reveals similar overall distributions from LIS and NaCl extractions.

profiles are shown in the upper panel. Black bar regions in the middle and lower panels identify S/MARs relative to the orange genes. In gene-dense regions (middle, the 30–31 Mb region), where many SARs are identified, regions of nuclear matrix attachment are often absent. In contrast, in gene-poor regions (lower, the 60–61 Mb region), where genes are often silenced, NaCl extraction reveals a multitude of nuclear matrix binding sites, whereas SARs are less prevalent.

To assess the local effect of S/MARs on transcript levels, the LIS and NaCl binding profiles from all attachments within as well as immediately 5′ and 3′ ends of all genes were analyzed by chi-square at 95% CI (1 *df*) with Yate's correction for continuity. This estimated the extent to which the measured parameters of S/MAR presence or absence at each 5′end and 3′ end region could also predict transcript presence. As summarized in Figure 5, correlative transcript level differences with SARs and MARs were revealed. A 5′ end SAR located at a distance of up to 10 kb from a gene correlates with expression of that gene, unlike MARs that show no significant correlation when located upstream of a gene. In contrast, the presence of a MAR within a gene correlates with a lack of transcripts, consistent with nuclear matrix attachment-induced silencing. Together, this data suggests that the different types of attachment, as revealed by either LIS or NaCl extraction, work in concert with other factors to modulate expression.

DISCUSSION

Recently, Wang *et al*. (2008) identified a histone modification module consisting of a combination of 17 modifications that are overrepresented at the promoters of genes that tend to be highly expressed in human CD4[+] T cells (38). However, it was cautiously noted that although the genes associated with the module tend to have higher expression, the histone modifications do not uniquely determine gene expression. This supports the notion that the modifications may not be the sole nucleating event in the chromatin remodeling that is seen in preparation for gene expression. There may be other events that coordinately act to regulate structure and poise a gene/locus for transcription including chromosomal looping and scaffold/matrix attachment (6). The distribution of S/MARs identified in this study relative to genes, and the differences observed between those isolated by the different extraction methods, suggests that these interactions can be defined and their effects predicted.

We have shown that gene-dense regions tend to loop away from a 2 M NaCl-prepared nuclear matrix and that this is accentuated at the telomeric regions. In comparison, the AT-rich, gene-poor, G-banded regions are matrix enriched. This is consistent with previous observations of an inverse correlation of matrix attachment with gene density (39) as well as recent genomic analysis of MARs at the human MHC locus (40). This general pattern of organization is reminiscent of gene ridges (41) and open/closed chromatin fibers (42). It appears that at least three classes of interaction consistent with functional classes of attachment have been resolved (8). MARs appeared as peaks of enrichment that were biased towards intergenic regions, whereas SARs exhibited a more even distribution across the chromosome as expected from the global profile of scaffold enrichment (Fig. 1). The subset of SAR and MAR regions that were in common was not biased towards either genic or intergenic regions.

It is expected that restriction digestion will preferentially remove the apex of each loop, leaving the S/MAR attached along with immediately proximal DNA. Estimation of loop size based on the spacing between neighboring MARs and SARs revealed that the majority of MAR and SAR spacing
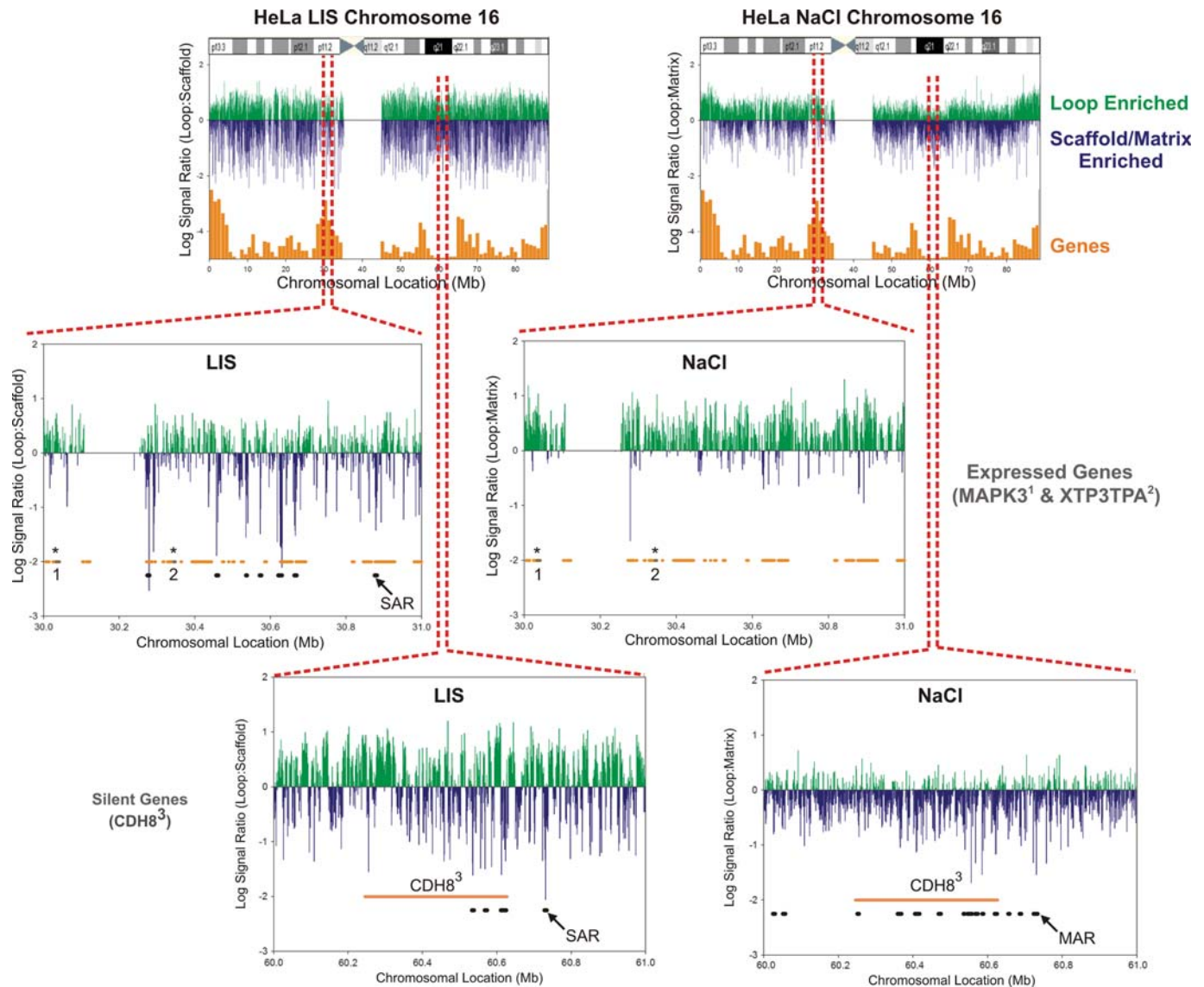
Figure 4. LIS and NaCl extractions reveal different profiles relative to gene density/expression. The nuclear scaffold/matrix aCGH profiles in a 30–31 Mb, 16p11.2 gene-rich and 60–61 Mb, 16q21 gene-poor region are shown Log$_2$ loop enrichment is indicated in green and scaffold/matrix association in blue. Regions identified as S/MARs (black bars) were compared with orange: silent and gray: expressed genes (indicated by asterisks). Gene-dense regions show little, if any nuclear matrix attachment and rather are bound to a LIS extracted nuclear scaffold. Gene-poor regions show significant nuclear matrix attachment with fewer-attachment sites to the LIS-isolated nuclear scaffold.

is similar to the 86 kb loop size identified using a physiological extraction protocol (25). Of particular interest is the subset of both SARs and MARs that are clustered to create loops in the range of 87 to 2217 bp. The median restriction fragment size is ~4.6 kb. However, the sizes of restriction fragments range from <100 bp to >10 kb consistent with the limits observed. Although the subset of small loop domains appears to be evenly distributed along the entire length of each chromosome, there are several regions on the chromosome where two small loops are separated by a single large loop. This organization is consistent with the clustering of S/MARs at the boundaries of a single large loop that would effectively isolate the components of that loop from neighboring DNA.

The validation of some, but not all regions of attachment at 16q21 to a LIS-extracted HeLa cell nuclear scaffold that had previously been identified (19) requires consideration. Interestingly, loss of heterozygosity of chromosome 16 at q21 is a genomic characteristic of many breast cancers. Differences in attachment could indicate inherent instability of this region. SKY analysis of HeLa cells used in this study and aCGH hybridization comparing these cells to a normal human female reference genome (Promega, Madison, WI, USA) revealed chromosome 16 aneuploidy (data not shown) and showed that other chromosomes exhibited aneuploidy and recombination. In accord with the instability of the genome in cell lines that have been extensively passaged (43), the cells are still viable, yet the karyotype differs (44).

Significance of S/MAR Association with Gene Expression

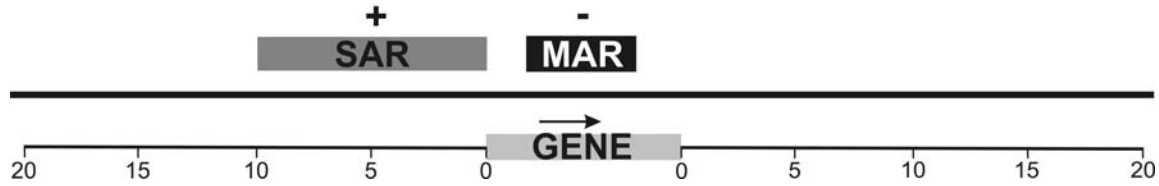|	20kb	10kb	5kb	2kb	1kb	Within Gene	1kb	2kb	5kb	10kb	20kb
SAR	0.197	**0.008**	**<.001**	**0.027**	**0.013**	0.736	1	0.777	0.519	0.914	0.325
MAR	0.987	0.848	0.356	0.729	0.859	**<.001**	0.139	0.147	0.954	0.97	0.89

Figure 5. SAR and MAR correlations with gene expression. The presence of S/MARs encompassing 20 kb region 5′ and 3′ of a gene, as well as within genes was compared with transcript presence using chi-square analysis (95% CI, 1 *df* with Yate's correction). The resulting *P*-values are displayed for each measurement with significance <0.05 indicated in bold. A SAR within 10 kb of the 5′ end region of a gene significantly correlates with gene expression. In contrast a MAR within a gene correlates with silencing of that gene. Direction of transcription is indicated by the arrow.

Together, these observations support the view that genome-wide instability resolving as a cell line/culture-specific spatial disruption of the nuclear scaffold/matrix presents as differences in their locations when various studies are compared. This tenet will be addressed when the nuclear scaffold/matrix binding profiles in primary cell lines are determined where genomic instability and prolonged cellular life in long-term cultures are not factors.

The presence of nuclear lamins supporting the nuclear envelope as well as dispersed throughout the nucleus suggests that these proteins may play a major structural role as a part of the nuclear scaffold/matrix. For example, the presence of lamin B1 throughout the nucleus as a part of the nucleoskeleton was recently shown to be necessary for RNA synthesis (45). The specific overlaps observed in this study between S/MARs and lamin B1-associated domains (36) may provide insight into the role of lamin proteins in the expression of our genome. However, the average length of the 37 LADs on chromosome 16 is 1.26 Mb, impeding the direct interpretation of the mechanistic significance.

We have shown that SARs located 5′ end of a gene, within and extending through the proximal promoter region, correlate with gene expression and may have a profound effect on whether a gene is transcribed. In contrast, MARs are generally located in gene-poor regions and at larger distances from expressed genes than SARs. However, the subset of MARs located within genes correlates with their silencing. This suggests that SARs may spatially poise a region of the genome for transcription and/or recruit factors necessary for genomic remodeling in preparation for transcription while attachment to a nuclear matrix as revealed by NaCl extraction may provide a means to restrict transcription.

The elucidation of the chromosome-wide distribution of S/MARs and their correlation with gene expression in HeLa cells has suggested a model of organizational architecture in which SARs and MARs are complementary predictors of whether a gene lies in a silenced or potentiated chromatin state. This supports a model of organization that functions with other architectural elements to bring regions of the genome into intimate contact with the factors that control expression. It is clear that at least structurally, attachment to the nuclear scaffold/matrix contributes to the modulation of gene expression. The S/MARs biologically delineated in this study begin to provide the extended sequence evidence that has frequently been called for and until now, not been available to develop a robust *in silico* model.

MATERIALS AND METHODS

Isolation and purification of S/MARs and loop regions

S/MARs and loop regions were prepared by either NaCl (46) or LIS (47) extraction. The optimal extraction time to remove histones and non-matrix nuclear proteins with 2 M NaCl was first determined as described (46). Nuclear halos were then prepared in solution from isolated HeLa nuclei with either the optimal timed exposure to 2 M NaCl as determined earlier or by dounce homogenization in the presence of 25 mM LIS as described (47) using $\sim 1 \times 10^7$ cells. After extraction, the halos were pelleted at $1000\,g$ for 5 min at 4°C then washed gently in REact® 3 restriction buffer (Invitrogen, Carlsbad, CA, USA) on a rocker platform for 20 min at room temperature then centrifuged at $1000\,g$ for 5 min at 4°C then the supernatant discarded. This washing procedure was repeated an additional two times. After the third wash, the halos were resuspended in restriction buffer and the loops separated from the nuclear matrix/scaffold-associated DNA by digestion with 400 U of *Eco*RI (Invitrogen) at 37°C for 3 h. Subsequent to restriction digestion, the matrix/scaffold fractions were pelleted by centrifugation at 16 000 g for 5 min at 4°C and the loop containing supernatants were removed and placed in separate tubes. The matrix/scaffold fractions were resuspended, and then washed in restriction buffer, immediately pelleted at 16 000 g for 5 min at 4°C and supernatant discarded. The nuclear matrix/scaffold containing pellet fraction was washed an additional two times. Both loop and matrix/scaffold restriction fragments were then freed from any nuclear proteins by overnight digestion at 55°C with 50 μg/ml of proteinase K buffered with 50 mM Tris–HCl buffer, pH 8.0, containing 50 mM NaCl, 25 mM EDTA and 0.5% SDS. DNA was recovered and purified using a Quantum-prep matrix (BioRad, Hercules, CA, USA) then resuspended in deionized water.

Verification of fractionation and CGH hybridization

The separation of loop and matrix/scaffold DNA was assessed by real-time PCR analysis as described (16). Regions previously shown to be matrix-associated or loop-enriched, including the human β-globin HS3 (16) and protamine 2 (48) regions, respectively, were initially amplified in triplicate from each fraction. Upon verification of expected fractionation, the remaining portions of the samples were utilized to identify loop-enriched and matrix/scaffold-attached fragments along the chromosomes. Purified DNA from each nuclear matrix/scaffold and loop fraction was analyzed using the array containing human chromosomes 14, 15, 16, 17 and 18 (Array 7 of the 8 array set) from the Nimblegen Systems CGAR0150-WHG8 CGH isothermal oligonucleotide array system (Nimblegen Systems, Inc., Madison, WI, USA). These arrays offer a median probe spacing of 713 bp, however, the experiments are limited to the ∼4.6 kb median length of EcoR1 restriction fragments. All experiments were replicated in independent preparations.

Identification and confirmation of S/MARs

CGH array data was initially viewed using Nimblescan version 1.9 (Nimblegen Systems, Inc.) for broad validation of similarity between replicates. Dual-channel array data were q spline normalized $\log_2$ using the NimbleGen data analysis suite. The normalized ratios of the loop and matrix/scaffold signals were not found to be simple symmetric distributions that could be readily transformed for parametric statistics (i.e. Kolmovorov–Smirnov test for normality failed for all datasets). Hence, rank comparisons were undertaken to assess concordance between biological replicates and nonparametric statistics were employed to establish peaks of enrichment.

Initial PCR validation of the array data suggested that similar signal distribution of additional nearby probes to the signal peaks was able to denote real signal and thus eliminate the effect of false hybridization to single probes and minimize the false-positive rate. Furthermore, the binning of probe signals was found to increase the correlation between replicates in a non-linear manner whereby a significant increase in correlation coefficients is observed up to a bin of 3 kb but levels off thereafter. One might expect that correlation would increase until the point at which the average restriction fragment size is met, as is the case here. By first analyzing the data solely on the basis of neighboring probe similarity as opposed to signal averages within EcoR1 fragments, the effect of outliers within restriction fragments that contain only a single array probe is ablated. Therefore, the regions of significance were identified initially as those probes with a signal ratio in the top 2.5% of the ranked signal. Probes located within 3 kb on either side of the top probes were then analyzed for similar signal distribution with a minimum requirement that at least two were present for inclusion of the region. Restriction fragments containing probes meeting these criteria were then analyzed for consistent signal across the entire fragment. Upon validation of consistent signal for each replicate, consistency was validated by comparing the two independent biological replicates. Regions identified in

this manner by both replicate experiments are presented. All S/MAR locations identified are available in Supplementary Material, Table S1. Loop regions for significance comparisons were identified in an analogous manner and are available in Supplementary Material, Table S2.

Eighteen regions of chromosome 16 were randomly selected to ensure an unbiased representation of loop or scaffold/matrix regions for real-time PCR verification as described (16). These regions represented both genic and intergenic segments across the chromosome and included a sampling of both loop-enriched regions and S/MARs as identified by aCGH. All PCRs were performed in triplicate starting with the same concentration of loop or matrix DNA. Initial template was calculated by the KLab PCR algorithm and ratios compared to array data as described (32). The percent enrichment of either loop or matrix relative to the total loop plus matrix template was then calculated. This was compared with the analogous percent total calculated from the independent loop or matrix array signal channel relative to the sum of the loop and matrix signals. For both datasets there was significant concordance between array identification and PCR validation. Regions that showed discordance between the array data and PCR validation displayed $\log_2$ signal ratios near zero, indicating that they are approximately equal parts loop and matrix. All primer sequences and ratios are available in Supplementary Material, Table S3.

Expression analysis

The expression profile of the HeLa cells used for aCGH analysis was determined. Total RNA was isolated using RNeasy (Qiagen Inc., Valencia, CA, USA). The RNA was then amplified using the Illumina RNA amplification system (Ambion, Austin, TX, USA) and 750 ng was used for hybridization to Illumina Sentrix Human-8 v2 Expression BeadChip arrays. Data was analyzed using the Illumina Bead Studio software suite. The average signal for each reporter was cubic spline normalized between chips to derive a standardized expression value. Expressed genes were identified by signal values higher than internal spike-in controls for expression ($S_{min} > 3000$).

Data analysis

Correlation of S/MARs between extraction methods as well as with expressed and silent genes was carried out using several tools including Suite 16 (49) and RegionMiner (Genomatix Software GmbH). When trends were detected, statistical significance was assessed using Sigma Stat (http://www.systat.com) and SPSS (http://www.spss.com). Chi-square analysis was conducted using Sigma Stat to establish the significance of any relationship between a propensity for genes to be expressed and the scaffold/matrix binding evidenced around them. Gene expression status was assigned a value of 1, where maximum expression over the possible reporters for each gene was expressed ($S_{min} > 3000$). All other genes for which expression was below this threshold level were assigned a value of zero. The scaffold/matrix association state of DNA between gene model 5′ end and 3′ end limits was used as the first variable where detection of one or more S/MARs assigned this variable a value of 1 (scaffold/matrix-associated) otherwise

0. Further variables were assigned in the same way from the scaffold/matrix association state of proximate regions beyond each gene over 1, 2, 5, 10 and 20 kb spans 5′ and 3′ of the Refseq gene model. Necessarily some of the scaffold/matrix-associated regions were associated with multiple genes and no requirement was introduced that they be either uniquely assigned to one gene or intergenic with respect to neighboring genes. All RefSeq genes covered by both the expression arrays and the aCGH arrays were assigned expression states and scaffold/matrix association states. Each combination of expression and location was compared by chi-square analysis at 95% CI with 1 *df* followed by Yate's correction for continuity in order to generate the *P*-values for each independent variable. Extractions were then each assessed independently.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Cremer, T. and Cremer, C. (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.*, **2**, 292–301.
2. Volpi, E.V., Chevret, E., Jones, T., Vatcheva, R., Williamson, J., Beck, S., Campbell, R.D., Goldsworthy, M., Powis, S.H., Ragoussis, J. *et al.* (2000) Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J. Cell Sci.*, **113** (Pt), 1565–1576.
3. Mahy, N.L., Perry, P.E. and Bickmore, W.A. (2002) Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *J. Cell Biol.*, **159**, 753–763.
4. Albiez, H., Cremer, M., Tiberi, C., Vecchio, L., Schermelleh, L., Dittrich, S., Kupper, K., Joffe, B., Thormeyer, T., von Hase, J. *et al.* (2006) Chromatin domains and the interchromatin compartment form structurally defined and functionally interacting nuclear networks. *Chromosome Res.*, **14**, 707–733.
5. Bode, J., Benham, C., Knopp, A. and Mielke, C. (2000) Transcriptional augmentation: modulation of gene expression by scaffold/matrix-attached regions (S/MAR elements). *Crit. Rev. Eukaryot. Gene Expr.*, **10**, 73–90.
6. Martins, R.P. and Krawetz, S.A. (2007) Decondensing the protamine domain for transcription. *Proc. Natl. Acad. Sci. USA*, **104**, 8340–8345.
7. Bode, J., Winkelmann, S., Gotze, S., Spiker, S., Tsutsui, K., Bi, C., A, K.P. and Benham, C. (2006) Correlations between scaffold/matrix attachment region (S/MAR) binding activity and DNA duplex destabilization energy. *J. Mol. Biol.*, **358**, 597–613.
8. Kramer, J.A. and Krawetz, S.A. (1996) Nuclear matrix interactions within the sperm genome. *J. Biol. Chem.*, **271**, 11619–11622.
9. Ottaviani, D., Lever, E., Takousis, P. and Sheer, D. (2008) Anchoring the genome. *Genome Biol.*, **9**, 201.
10. Martins, R.P., Ostermeier, G.C. and Krawetz, S.A. (2004) Nuclear matrix interactions at the human protamine domain: a working model of potentiation. *J. Biol. Chem.*, **279**, 51862–51868.
11. Farrell, C.M., West, A.G. and Felsenfeld, G. (2002) Conserved CTCF insulator elements flank the mouse and human beta-globin loci. *Mol. Cell Biol.*, **22**, 3820–3831.
12. Goetze, S., Baer, A., Winkelmann, S., Nehlsen, K., Seibler, J., Maass, K. and Bode, J. (2005) Performance of genomic bordering elements at predefined genomic loci. *Mol. Cell Biol.*, **25**, 2260–2272.
13. Yusufzai, T.M. and Felsenfeld, G. (2004) The 5′-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element. *Proc. Natl. Acad. Sci. USA*, **101**, 8620–8624.
14. Mesner, L.D., Hamlin, J.L. and Dijkwel, P.A. (2003) The matrix attachment region in the Chinese hamster dihydrofolate reductase origin of replication may be required for local chromatid separation. *Proc. Natl. Acad. Sci. USA*, **100**, 3281–3286.
15. Debatisse, M., Toledo, F. and Anglana, M. (2004) Replication initiation in mammalian cells: changing preferences. *Cell Cycle*, **3**, 19–21.
16. Ostermeier, G.C., Liu, Z., Martins, R.P., Bharadwaj, R.R., Ellis, J., Draghici, S. and Krawetz, S.A. (2003) Nuclear matrix association of the human beta-globin locus utilizing a novel approach to quantitative real-time PCR. *Nucleic Acids Res.*, **31**, 3257–3266.
17. Cai, S., Lee, C.C. and Kohwi-Shigematsu, T. (2006) SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat. Genet.*, **38**, 1278–1288.
18. Ioudinkova, E., Petrov, A., Razin, S.V. and Vassetzky, Y.S. (2005) Mapping long-range chromatin organization within the chicken alpha-globin gene domain using oligonucleotide DNA arrays. *Genomics*, **85**, 143–151.
19. Shaposhnikov, S.A., Akopov, S.B., Chernov, I.P., Thomsen, P.D., Joergensen, C., Collins, A.R., Frengen, E. and Nikolaev, L.G. (2007) A map of nuclear matrix attachment regions within the breast cancer loss-of-heterozygosity region on human chromosome 16q22.1. *Genomics*, **89**, 354–361.
20. Purbowasito, W., Suda, C., Yokomine, T., Zubair, M., Sado, T., Tsutsui, K. and Sasaki, H. (2004) Large-scale identification and mapping of nuclear matrix-attachment regions in the distal imprinted domain of mouse chromosome 7. *DNA Res.*, **11**, 391–407.
21. Jackson, D.A., McCready, S.J. and Cook, P.R. (1984) Replication and transcription depend on attachment of DNA to the nuclear cage. *J. Cell Sci. Suppl.*, **1**, 59–79.
22. Vogelstein, B., Pardoll, D.M. and Coffey, D.S. (1980) Supercoiled loops and eucaryotic DNA replicaton. *Cell*, **22**, 79–85.
23. Heng, H.H., Goetze, S., Ye, C.J., Liu, G., Stevens, J.B., Bremer, S.W., Wykes, S.M., Bode, J. and Krawetz, S.A. (2004) Chromatin loops are selectively anchored using scaffold/matrix-attachment regions. *J. Cell Sci.*, **117**, 999–1008.
24. Heng, H.H., Krawetz, S.A., Lu, W., Bremer, S., Liu, G. and Ye, C.J. (2001) Re-defining the chromatin loop domain. *Cytogenet. Cell Genet.*, **93**, 155–161.
25. Jackson, D.A., Dickinson, P. and Cook, P.R. (1990) The size of chromatin loops in HeLa cells. *Embo J.*, **9**, 567–571.
26. Platts, A.E., Quayle, A.K. and Krawetz, S.A. (2006) *In-silico* prediction and observations of nuclear matrix attachment. *Cell. Mol. Biol. Lett.*, **11**, 191–213.
27. Bode, J., Schlake, T., Rios-Ramirez, M., Mielke, C., Stengert, M., Kay, V. and Klehr-Wirth, D. (1995) Scaffold/matrix-attached regions: structural properties creating transcriptionally active loci. *Int. Rev. Cytol.*, **162A**, 389–454.
28. Ma, H., Siegel, A.J. and Berezney, R. (1999) Association of chromosome territories with the nuclear matrix. Disruption of human chromosome territories correlates with the release of a subset of nuclear matrix proteins. *J. Cell Biol.*, **146**, 531–542.
29. Goetze, S., Gluch, A., Benham, C. and Bode, J. (2003) Computational and in vitro analysis of destabilized DNA regions in the interferon gene

cluster: potential of predicting functional gene domains. *Biochemistry*, **42**, 154–166.

30. Donev, R.M. (2000) The type of DNA attachment sites recovered from nuclear matrix depends on isolation procedure used. *Mol. Cell Biochem.*, **214**, 103–110.

31. Linnemann, A.K., Platts, A.E., Doggett, N., Gluch, A., Bode, J. and Krawetz, S.A. (2007) Genomewide identification of nuclear matrix attachment regions: an analysis of methods. *Biochem. Soc. Trans.*, **35**, 612–617.

32. Platts, A.E., Johnson, G.D., Linnemann, A.K. and Krawetz, S.A. (2008) Real-time PCR quantification using a variable reaction efficiency model. *Anal. Biochem.*, **380**, 315–322.

33. Branco, M.R. and Pombo, A. (2006) Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol.*, **4**, e138.

34. Steinert, P.M. and Roop, D.R. (1988) Molecular and cellular biology of intermediate filaments. *Annu. Rev. Biochem.*, **57**, 593–625.

35. Luderus, M.E., de Graaf, A., Mattia, E., den Blaauwen, J.L., Grande, M.A., de Jong, L. and van Driel, R. (1992) Binding of matrix attachment regions to lamin B1. *Cell*, **70**, 949–959.

36. Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W. *et al.* (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, **453**, 948–951.

37. Poljak, L., Seum, C., Mattioni, T. and Laemmli, U.K. (1994) SARs stimulate but do not confer position independent gene expression. *Nucleic Acids Res.*, **22**, 4386–4394.

38. Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q. *et al.* (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.*, **40**, 897–903.

39. Craig, J.M., Boyle, S., Perry, P. and Bickmore, W.A. (1997) Scaffold attachments within the human genome. *J. Cell Sci.*, **110** (Pt), 2673–2682.

40. Ottaviani, D., Lever, E., Mitter, R., Jones, T., Forshew, T., Christova, R., Tomazou, E.M., Rakyan, V.K., Krawetz, S.A., Platts, A.E. *et al.* (2008) Reconfiguration of genomic anchors upon transcriptional activation of

the human major histocompatibility complex. *Genome Res.*, **18**, 1778–1786.

41. Goetze, S., Mateos-Langerak, J., Gierman, H.J., de Leeuw, W., Giromus, O., Indemans, M.H., Koster, J., Ondrej, V., Versteeg, R. and van Driel, R. (2007) The three-dimensional structure of human interphase chromosomes is related to the transcriptome map. *Mol. Cell Biol.*, **27**, 4475–4487.

42. Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P. and Bickmore, W.A. (2004) Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell*, **118**, 555–566.

43. Bazer, F.W. and Salamonsen, L.A. (2008) Let's validate those cell lines. *Biol. Reprod.*, **79**, 585.

44. Macville, M., Schrock, E., Padilla-Nash, H., Keck, C., Ghadimi, B.M., Zimonjic, D., Popescu, N. and Ried, T. (1999) Comprehensive and definitive molecular cytogenetic characterization of HeLa cells by spectral karyotyping. *Cancer Res.*, **59**, 141–150.

45. Tang, C.W., Maya-Mendoza, A., Martin, C., Zeng, K., Chen, S., Feret, D., Wilson, S.A. and Jackson, D.A. (2008) The integrity of a lamin-B1-dependent nucleoskeleton is a fundamental determinant of RNA synthesis in human cells. *J. Cell Sci.*, **121**, 1014–1024.

46. Krawetz, S.A., Draghici, S., Goodrich, R., Liu, Z. and Ostermeier, G.C. (2005) *In silico* and wet-bench identification of nuclear matrix attachment regions. *Methods Mol. Med.*, **108**, 439–458.

47. Kay, V. and Bode, J. (1995) Detection of scaffold-attached regions (SARs) by in vitro techniques; activities of these elements *in vivo*. Papavassiliou, A.G. and King, S.L. (eds), Methods in Molecular and Cellular Biology: Methods for studying DNA–protein interactions—An Overview.Wiley-Liss, Vol. **5**, pp. 186–194.

48. Kramer, J.A., Adams, M.D., Singh, G.B., Doggett, N.A. and Krawetz, S.A. (1998) Extended analysis of the region encompassing the PRM1–>PRM2–>TNP2 domain: genomic organization, evolution and gene identification. *J. Exp. Zool.*, **282**, 245–253.

49. Naismith, L., Lalancette, C., Platts, A.E. and Krawetz, S.A. (2008) The KLAB Toolbox: A suite of in-house software applications for epigenetic analysis. *SBiRM: Syst. Biol. Reprod. Med.*, **54**, 97–108.