Distribution of Nucleosome-enriched Sequences of Human Sperm Chromatin Along Isochores

Stilianos Arhondakis¹ and Annalisa Varriale²*

¹Software Technology and Network Applications Lab, School of Electrical & Computer Engineering, Technical University of Crete, Crete, Greece; ²School of Medicine and Surgery, Università degli Studi di Napoli Federico II, Naples, Italy

Abstract

Background and objectives: During sperm maturation, chromatin undergoes an extensive compaction through the progressive replacement of histones with basic, positively-charged proteins called protamines. Interestingly, in human and other vertebrates, 5–15% of the paternal genome retains a somatic-like structure. It has been suggested that preserved nucleosomes have a role in promoting transcription after fertilization, however their localization and function is still a matter of debate. The aim of the present work was to understand if the localization of histones in human sperm DNA is affected by genome composition, a feature linked to several biological functions. In order to do that, we mapped histone-enriched sequences along isochores, which are large chromosomal territories characterized by fairly homogeneous guanine and cytosine (GC) composition.

Methods: We retrieved publicly available human sequences found to be histone-enriched in sperm chromatin and localized them along isochores by using their respective coordinates.

Results: We found that the majority of genes and sequences associated with mononucleosomes, the activating chromatin mark H3K4me3 and the repressive mark H3K27me3 reside in the GC-poor isochores. Genes harboring the activating mark H3K4me2, instead, are preferentially located in the GC-rich isochores.

Conclusions: For genes carrying histones with specific marks, we speculate that developmental and tissue-specific genes (carrying the H3K27me3 mark) might take advantage by residing in GC-poor isochores because in these regions they can be shut off immediately after the end of their transcription more easily than in GC-rich isochores. On the contrary, housekeeping genes (marked with H3K4me2) would not need this kind of regulation and reside in GC-rich isochores. For other genes and sequences, further analyses are required in order to clarify the role of GC composition in histone retention in sperm chromatin. In conclusion, although the complete scenario of sperm chromatin has not been elucidated yet, correlations seem to exist between chromatin modification patterns, isochore structure and gene expression timing in embryo development.

Keywords: Sperm chromatin; Vertebrates; Isochores; Nucleosomes; Early expression.

Abbreviation: GC, guanine and cytosine.

Received: May 24, 2018; Revised: August 04, 2018; Accepted: September 13, 2018

*Correspondence to: Annalisa Varriale, School of Medicine and Surgery, Università degli Studi di Napoli Federico II, Naples, Italy. E-mail: avarrial@libero.it

How to cite this article: Arhondakis S, Varriale A. Distribution of Nucleosome-enriched Sequences of Human Sperm Chromatin Along Isochores. Exploratory Research and Hypothesis in Medicine 2018;3(3):54–60. doi: 10.14218/ERHM.2018.00009.

Exploratory Research and Hypothesis in Medicine 2018 vol. 3 | 54–60

Copyright: © 2018 Authors. This is an Open Access article distributed under the terms of the Creative Commons Attribution-Noncommercial 4.0 International License (CC BY-NC 4.0), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Arhondakis S. et al.: Distribution of nucleosome-enriched sequences total DNA and bound to the preserved histones. It was hypothesized that histone retention was nonrandom and had a role in ensuring proper gene expression, correct development, totipotency and inheritance of DNA methylation in the early embryo.  

Among the first findings that supported the nonrandomness of histone retention, two studies on the β-globin cluster demonstrated that the early expressed isoforms and those expressed in the adult were associated to histones and protamines, respectively. Subsequently, it was demonstrated that histone-bound DNA was mainly enriched in: i) exons, in comparison with flanking introns; ii) gene-dense regions and developmentally-regulated promoters; iii) development-related gene families; iv) transcription start sites and CpG-rich promoters. In apparent contrast, other analyses put under discussion these results because the authors claimed that the majority of preserved nucleosomes was associated to repeat elements, such as LINEs and SINEs, pericentric repeats and GC-poor, gene-poor regions. Elsewhere, it was suggested that these conflicting findings may have resulted from technical issues. As a consequence, it has not yet been possible to reach a final conclusion about the role and localization of histones in sperm chromatin, and the debate is still ongoing.

In this study, our main aim was to clarify if the retained histones are preferentially distributed in the GC-rich or in the GC-poor regions of the human genome and if this association has a functional meaning. To do that, we calculated the frequencies of sequences and genes associated to mononucleosomes or carrying the epigenetic marks H3K27me3 (repressive mark), H3K4me3 and H3K4me2 (activating marks) in human isochores. The isochores are defined as Megabase-size regions of DNA (> 0.3 Mb, up to several Mb) fairly homogeneous in base composition and associated with a number of biological properties, such as gene expression and distribution, DNA methylation, CpG islands, replication timing and chromatin structure (as reviewed in Bernardi).

According to the isochrome map of human chromosomes, the total number of isochrome bands in the human genome is 3,159, a number close to the maximum number obtained for chromosomal bands at the highest resolution in early prophase. In vertebrate genomes, isochores are grouped into the following five families characterized by different ranges of GC level: two GC-rich families, represented by H3 (>53% GC) and H2 (ranging from 46–53% GC); and, three GC-poor families, represented by H1 (ranging from 41–46% GC), L2 (from 37–41% GC) and L1 (<37% GC). The GC-poor families, while representing approximately 85% of the genome, constitute the “genome desert” because they are characterized by a low gene density. The remaining 15%, represented by the three GC-rich families, is called the “genome core” since it contains more than half of the genes.

Given that both the retained histones and the GC-rich isochores represent approximately 15% of the genome, we hypothesized that their localization might coincide. If this hypothesis had been right, then we would have to find the totality of histones embedded in the GC-rich isochores. Unlike our expectations, we found that histones are distributed in large amounts also in the GC-poor isochores. At this point, we tried to find an explanation by examining the distribution of genes carrying epigenetic marks. We found that genes whose function is not limited to the first stages of development are mainly associated with GC-rich isochores, while developmental and tissue-specific genes are associated with the GC-poor ones. Our approach, taking into account isochores structure, provides—for the first time—information about the association between isochores and histone retention in human sperm chromatin, supporting the hypothesis that the distribution of nucleosomes in sperm chromatin is not random and might be related to expression timing in the early phases of development.  

Materials and methods

Retrieval of coordinates of selected sequences from the datasets of Hammoud et al. and Samans et al.

Accession numbers of genes enriched in mononucleosomes and chromatin marks in human sperm chromatin were downloaded from the Supplementary Material of Hammoud et al. In order to extract gene name, chromosome number, and transcription start and end sites for each given accession we used the Table Browser tool at Genome Browser (www.genome.ucsc.edu) with the following selection criteria: Assembly: hg17 (to allow compatibility with the coordinates of isochores); Group: Gene and Gene predictions track; Track: RefSeq Genes; Table: RefGene; Region: Genome. Entries with no match in RefGene were discarded. Subsequently, output was inspected manually in order to remove entries having the same or overlapping coordinates. As a result, for mononucleosomes, from an initial dataset of 654 entries, we obtained a final dataset of 522 sequences. For genes carrying the epigenetic marks H3K27me3 and H3K4me2, we obtained the coordinates and names of genes harboring epigenetic marks through the BIOMART tool at the Ensembl website (www.ensembl.org/info/data/biomart/index.html). Their positions along the isochores were determined as described in next paragraph. As control analysis, for these genes we also used the Table Browser tool, obtaining comparable results (data not shown).

To increase the strength of our results, we decided to also include in our analyses the sequences provided by Samans et al. since these authors put under discussion the findings of Hammoud et al. To do that, we downloaded information from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) the Supplementary file GSM1160359 containing the coordinates of 14,823 peaks, representing nucleosome binding sites in human sperm chromosomes. The coordinates were used to match these sequences to the isochrome bands. Finally, for each dataset, we pooled the data of all chromosomes and we calculated the frequency (in %), and the density of mononucleosomes and of chromatin marks (see below), dividing the number of genes by the length of isochores in Mb.

Matching histone-enriched genes and sequences to the isochores

Coordinates of the human isochores and their GC levels were retrieved from Costantini et al. The latter were used to generate, for each chromosome, a table containing all isochrome bands with their relative positions (nucleotide start/end) and color (we used the color code reported in the literature: H3-red; H2-orange; H1-yellow; L2-light blue; L1-dark blue). This table was used to match each gene to the corresponding isochrome band. Finally, we plotted the frequencies of genes associated to mononucleosomes and modified histones. We also calculated the mononucleosome density (with the formula: d = number of genes/number of isochrome bands) across the isochrome families, obtaining comparable results (data not shown). Finally, we calculated and plotted the densities of genes harboring modified histones over the Mb of each isochrome.

DOI: 10.14218/ERHM.2018.00009 | Volume 3 Issue 3, September 2018
family, using the formula $d = 10 \times \text{(number of genes/Mb of isochores)}$.

**Results**

**Distribution and density of mononucleosomes in the isochrome families and their comparison with data from Samans et al.**

We localized genes associated to unmodified mononucleosomes on the isochrome table, and estimated their number within each isochrome family, and finally after pooling data for chromosomes 1 to 22 we plotted their percent frequencies. Results are shown in Figure 1a. From this figure it is visible that nearly half of the genes are located in the L2 isochrome and a quarter in H1, whereas only the 20% reside in the genome core (isochores H2 and H3). The above results suggest that the three-quarters of the genes retaining histones are embedded in GC-poor isochores (H1 and L2), which constitute the gene desert.

It is useful to remember that the gene desert is not totally devoid of genes, but they are less dense in this compartment compared to the genome core. Looking at the histogram of isochrome bands, reported for reference, it is possible to see that the frequency trend of histone-enriched genes follows that of isochores. When we considered the density, we could observe that it undergoes a slight increase from GC-poor to GC-rich regions (Fig. 2a). In Figure 1b we report the results of the analyses of sequences retrieved from Samans et al.; in this case, the percent of sequences found in GC-poor isochores is even higher.

Thereby, it appears that in both cases the majority of histone-enriched sites are embedded in GC-poor isochores. This result, however, is more pronounced with the data from Samans’ group, where only 10% of the sites belong to GC-rich isochores. This difference could be attributed to the fact that sequences from Samans et al include nongenic regions as well. In fact, when we consider the density, we can see a bias towards L1 and L2 isochores, which most probably can be attributed to the presence of nongenic re-

![Fig. 1. Distribution of genes and sequences from literature across human isochrome families.](image)

(a) The histogram displays the percent frequencies of the genes associated to mononucleosomes found in sperm chromatin along the five isochrome families of human genome. L1, L2 and H1 represent the GC-poor isochores (around 80% of the total genome), whereas H2 and H3 represent the GC-rich compartment (around 20% of the total genome). Chromosomes 1 to 22, whose data were provided by Hammoud SS, et al were considered. (b) The histogram displays the percent frequencies of sequences associated to histones in sperm chromatin from Samans B, et al. In both panels, the coloured bars refer to the isochores and the colour code is that specified in literature (see text).

![Fig. 2. Plots showing density of sequences from literature across isochores.](image)

(a) The histogram displays the density (10 × number of genes/Mb of isochrome) of the genes associated to mononucleosomes. (b) The histogram displays the density (number of genes/Mb of isochrome) of the sequences associated to histones in sperm chromatin from Samans et al. In both cases, we reported the histogram of the percent frequency of isochrome families for reference (see Fig. 1).
Distribution of nucleosome-enriched sequences

We estimated the frequency (in %) of genes associated with the chromatin marks H3K27me3 (repressive), H3K4me2 and H3K4me3 (activating) across the isochores (Fig. 3a, b and c, respectively). Each column represents the frequency (in %) of genes carrying marked histones in each isochore family of human chromosomes. As described previously for Figure 1, the data for all chromosomes were pooled and then we reported the percent frequency of isochores for reference. As can be seen, genes associated with H3K27me3 are mostly (~70%) located in GC-poor isochores, in agreement with the results obtained by Jung et al who found that this modification is enriched in the gene-poor compartment. Concerning H3K4me3, the total number of genes associated to this modification is lower, but the distribution is similar to that of the repressive mark. In fact, it has been proposed that these two modifications are often associated, constituting a bivalent chromatin structure, whose function should be to repress genes maintaining them and able to be activated at the same time (gene poising). On the contrary, Figure 3 shows that genes harboring H3K4me2 are enriched in GC-rich isochores.

In Figure 4 we plotted the percent frequencies of the three modifications, together with that of the isochore bands. Remarkably, our analysis reveals an opposite trend between H3K4me2 and H3K27me3 in relation to the frequency of the isochore families. The frequency of H3K4me2, in fact, presents a strikingly opposite trend respective to that of the frequency of isochore bands, whereas the frequency of H3K27me3 and of H3K4me3 are almost equally distributed along the five isochore families. We calculated and plotted the density (number genes/Mb) of each of the above modified histones along the isochore families. We could see that the densities of the three modifications are correlated with the GC level of isochores; in fact, gene density is higher in the GC-rich compartment.

As depicted in Figure 5, the density of H3K4me3 is the highest in H3, probably because this mark is associated with housekeeping genes, which are more frequent in H3.

Fig. 3. Distribution of genes carrying histone modifications across human isochore families. (a) Distribution of the genes associated to H3K27me3 (grey bars) for each isochore family. (b) Distribution of the genes associated to H3K4me2 (grey bars) for each of the isochore family. (c) Distribution of H3K4me3 (grey bars). The frequencies of the isochore families are shown in all panels for reference.

Fig. 4. Plot showing percent frequency of modified histones across isochore families. Distribution of the genes associated to chromatin marks in human chromosomes in relation to isochores. Brown dotted line: H3K4me3; green full line: H3K27me3; purple dashed line: H3K4me2. The histogram represents the profile of the isochore families, reported as a reference.

Fig. 5. Plot showing densities of modified histones in the isochore families. Densities (number of genes/Mb of isochore) of the three chromatin marks across isochores in human chromosomes. Brown dotted line: H3K4me3; green full line: H3K27me3; purple dashed line: H3K4me2.
In this study, for the first time, genomic sequences that retain histones in human sperm chromatin have been mapped along the isochore families in order to correlate their distribution with the histones in human sperm chromatin have been mapped along the promoters and genes,\textsuperscript{24,30,33,34} while some studies suggest that histones are bound mostly to repetitive sequences.\textsuperscript{39,43} However, the association of nucleosomes with repetitive sequences could have a role in postfertilization processes reviewed in Castillo et al.,\textsuperscript{13} in tridimensional organization of male gamete chromatin and in inheritance of paternal chromatin structure.\textsuperscript{54}

Interestingly, regarding point i) above, if we compare the plot showing nucleosome retention with the isochore plot,\textsuperscript{12,24} we observe that most of the GC peaks associated with nucleosomes are actually located in GC-poor isochores. This scenario becomes very clear when we consider the case of HOX clusters that are all GC-rich sequences (which then appear as GC peaks) embedded into larger GC-poor isochores.

About point ii), Barton and coworkers showed that during the passage from early to late stages of preimplantation processes,\textsuperscript{55} GC-rich isochores increase their expression activity while GC-poor ones decrease it. Moreover, it has been reported that most genes expressed during mouse brain development are located in GC-poor and LINE-rich regions.\textsuperscript{56,57} On the basis of these lines of evidence, it has been proposed that genes located in GC-poor isochores are actively expressed during development and turn silent by heterochromatinization at the end of this process.\textsuperscript{58,59} In this context, we can speculate that the presence of bivalent chromatin (H3K27me3/H4K3me3) in GC-poor regions could help to facilitate this process.

### Conclusions

During the last years, the interest and progress towards sperm genetics and epigenetics has increased deeply, leading to a clarification that sperm-persisting histones are not simply remnants of an inefficient replacement process but contribute to the paternal information delivered to the offspring. Nevertheless, there is an ongoing debate regarding the genomic location of retained histones.\textsuperscript{12,24,34,37,38,52} Our work proposes a “background approach”, in which we focused on the genome as a whole, through the isochores; this approach allowed us to overcome the vision of histones associated with genes or nongenic sequences, as we considered regions much larger compared to single sequences. Providing an explanation for this debate was not our goal; however, on the basis of our results, we can conclude that the two visions might be compatible.

We show that in human sperm DNA, genes and sequences enriched in histones are more frequent in GC-poor isochores, with the exception of genes associated to the activating mark H4K3me2. It is known that H3K4me3 covers genes related to development, organogenesis and tissue specificity, whereas H3K4me2 is more frequently associated with housekeeping genes.\textsuperscript{24} This is consistent with the fact that tissue-specific and developmental genes are enriched in GC-poor isochores, whereas housekeeping genes are more frequent in GC-rich ones.\textsuperscript{39}

On the basis of these observations, we suggest that housekeeping genes, whose expression is not limited to the first stages of development, would prefer a GC-rich isochore localization. On the
contrary, tissue-specific and developmental genes whose expression is time-limited can take advantage of a structure constituted by bivalent nucleosomes and GC-poor isochores, where they can easily be switched off to a silent state immediately after their transcription. Histone retention is, then, likely to be nonrandom in the sperm genome and could possibly be correlated to the isochores in which they are embedded and, ultimately, to the period in which the genes have to be expressed.49

Perspectives

This work opens new perspectives on the importance of isochores structure in the localization of genomic determinants, such as nucleosome-enriched sequences in sperm chromatin. Mapping sequences associated to histones along the chromosomes can be useful in order to display an eventual clustering, to match them with the determinants of tridimensional structure, such as matrix attachment regions and LADs and, especially, to discover signals that might specify retention. It could also be interesting to unravel a possible correlation between isochores, chromatin marks, other epigenetic regulators (such as DNA methylation and attachment to the nuclear matrix) and gene expression in sperm chromatin of animal models. A future step will be to characterize in detail the genomic localization of the single genes according to their expression activity during spermatogenesis or during the early zygotic stages.

Another relevant point for evolutionary and comparative genomics studies would be to understand how sperm chromatin marks are distributed along the isochores in other animal models, such as zebrafish, that do not use protamines to compact sperm chromatin such as zebrafish, that do not use protamines to compact sperm chromatin.50 Studying animal models differing in the extent of nucleosomal compaction and in organization of isochores, could increase our understanding of sperm packaging from an evolutionary point of view.

Conflict of interest

The authors have no conflict of interests related to this publication.

Author contributions

AV conceived the research, produced, analyzed and interpreted data, and wrote the manuscript. SA participated in data interpretation and in correcting the manuscript.

References


DOI: 10.14218/ERHM.2018.00009 | Volume 3 Issue 3, September 2018

59


