Original Article

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

Stilianos Arhondakis[1] and Annalisa Varriale[2*]

[1]Software Technology and Network Applications Lab, School of Electrical & Computer Engineering, Technical University of Crete, Crete, Greece; [2]School 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.

Introduction

In order to accomplish the various steps of fertilization, sperm chromatin undergoes a remarkably striking condensation by sub-stituting histones firstly with transition proteins and secondly with protamines, a class of positively-charged proteins expressed only in male germ cells.[1–3] In the course of evolution, this transition has recurred sporadically and independently in several metazoan groups,[4] with protamines always present at the upper tips of the phylogenetic tree and the main types of basic proteins widely represented across both invertebrates, such as molluscs, tunicates, insects, and vertebrates.[5–12]

In human, during spermatogenesis approximately 90% of the canonical histones are replaced by protamines (reviewed in Castillo *et al*),[13] so that paternal chromatin results packaged into three major structural domains: one attached to the nuclear Matrix Attachment Regions (MARs)[14,15]; a second one containing the majority of sperm DNA, coiled into toroids by protamines[16]; and, a third one corresponding to approximately 5–15% of the

total DNA and bound to the preserved histones.[17–24] 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.[17,25–29]

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.[30,31] Subsequently, it was demonstrated that histone-bound DNA was mainly enriched in: i) exons, in comparison with flanking introns[32]; ii) gene-dense regions and developmentally-regulated promoters[33]; iii) development-related gene families[24,34]; iv) transcription start sites and CpG-rich promoters.[35]

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,[21] pericentric repeats[36] and GC-poor, gene-poor regions.[7,12] Elsewhere, it was suggested that these conflicting findings may have resulted from technical issues.[37] 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.[38]

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),[24] H3K4me3 and H3K4me2 (activating marks) in human isochores.[24] 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,[39] such as gene expression and distribution,[40] DNA methylation, CpG islands, replication timing and chromatin structure (as reviewed in Bernardi).[39]

According to the isochore map of human chromosomes, the total number of isochore 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.[41] In vertebrate genomes, isochores are grouped into the following five families characterized by different ranges of GC level[39]: 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.[41]

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 isochore 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[12,24]

Accession numbers of genes enriched in mononucleosomes and chromatin marks in human sperm chromatin were downloaded from the Supplementary Material of Hammoud *et al.*[24] 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)[41,42]; 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*,[12] since these authors put under discussion the findings of Hammoud *et al.*[24] 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,[12] representing nucleosome binding sites in human sperm chromosomes. The coordinates were used to match these sequences to the isochore 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.*[41] The latter were used to generate, for each chromosome, a table containing all isochore 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).[39,41–43] This table was used to match each gene to the corresponding isochore 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 isochore bands) across the isochore families, obtaining comparable results (data not shown). Finally, we calculated and plotted the densities of genes harboring modified histones over the Mb of each isochore
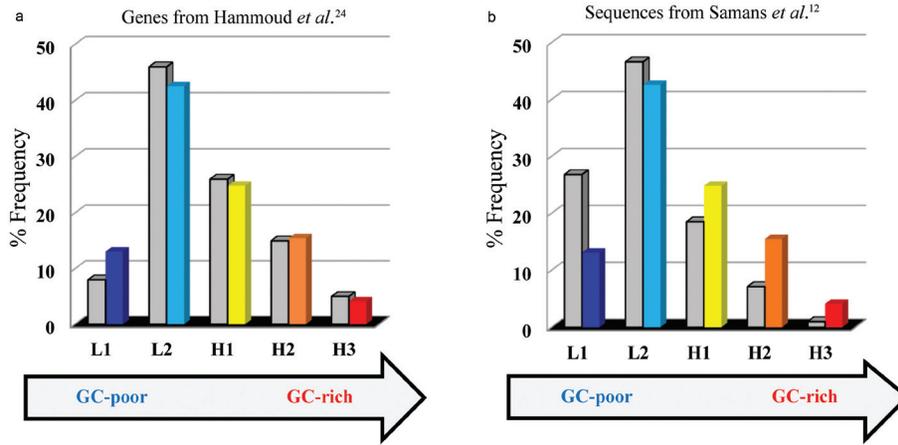
Fig. 1. Distribution of genes and sequences from literature across human isochore families. (a) The histogram displays the percent frequencies of the genes associated to mononucleosomes found in sperm chromatin along the five isochore families of human genome.[24] L1, L2 and H1 represent the GC-poor 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.[24] (b) The histogram displays the percent frequencies of sequences associated to histones in sperm chromatin from Samans B, *et al*.[12] In both panels, the coloured bars refer to the isochores and the colour code is that specified in literature (see text).

family, using the formula d = 10 × (number of genes/Mb of iso-chores).

Results

Distribution and density of mononucleosomes in the isochore families and their comparison with data from Samans et al[12,24]

We localized genes associated to unmodified mononucleosomes on the isochore table,[24] and estimated their number within each isochore 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 isochore 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.[39] Looking at the histogram of isochore 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*[12]; 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.[12,24] This result, however, is more pronounced with the data from Samans' group,[12] 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.[12] 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-
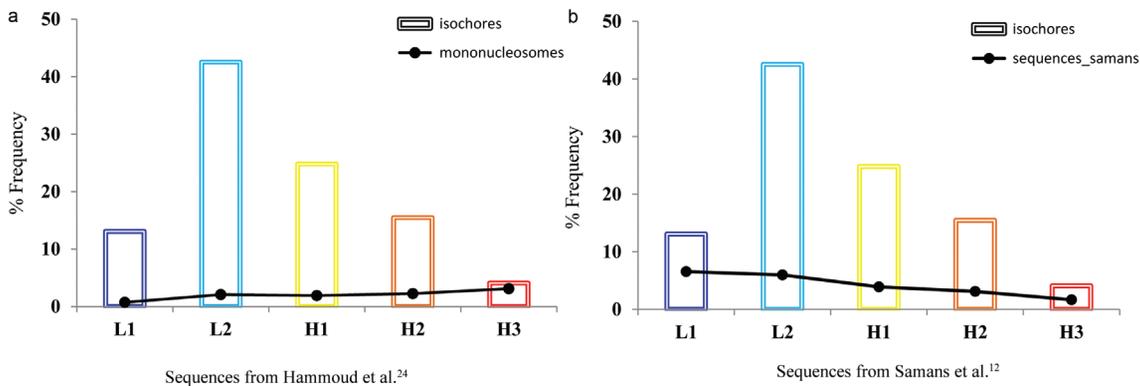
Fig. 2. Plots showing density of sequences from literature across isochores.[12,24] (a) The histogram displays the density (10 × number of genes/Mb of isochore) of the genes associated to mononucleosomes.[24] (b) The histogram displays the density (number of genes/Mb of isochore) of the sequences associated to histones in sperm chromatin from [12]. In both cases, we reported the histogram of the percent frequency of isochore families for reference (see Fig. 1).
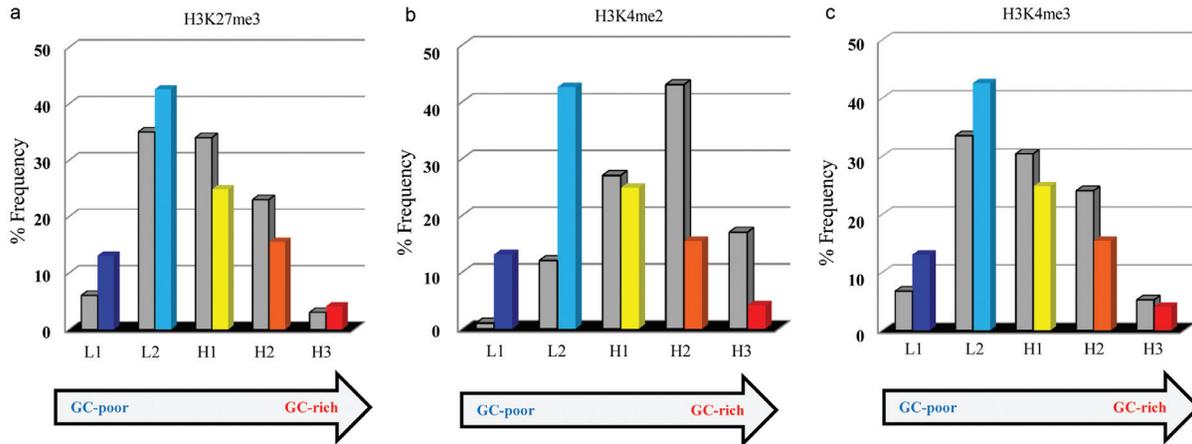
Fig. 3. Distribution of genes carrying histone modifications across human isochore families.[24] (a) Distribution of the genes associated to H3K27me3 (grey bars) for each isochore family. (b) Distribution of the genes associated to H3K4me2 (grey bars) for each of the isochore family. (c) Distribution of H3K4me3 (grey bars). The frequencies of the isochore families are shown in all panels for reference.

gions (Fig. 2b).

Distribution and density of H3K27me3, H3K4me3 and H3K4me2 in isochore families

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).[24] 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.[44] 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 associ-

ated, 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).[24] 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 (Fig. 5); in fact, gene density is higher in the GC-rich compartment. More precisely, this is evident for H3K4me2, the density of which is highest in H3, probably because this mark is associated with housekeeping genes, which are more frequent in H3.[39]
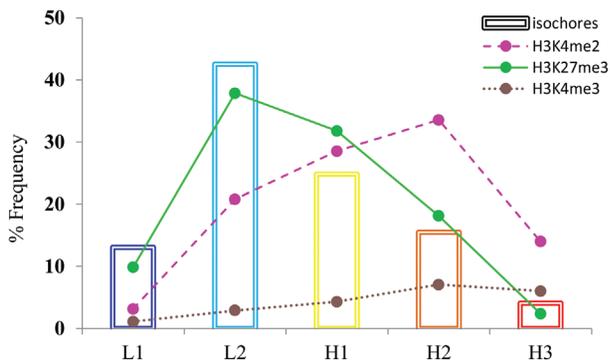
Fig. 4. Plot showing percent frequency of modified histones across isochore families. Distribution of the genes associated to chromatin marks in human chromosomes in relation to isochores. Brown dotted line: H3K4me3; green full line: H3K27me3; purple dashed line: H3K4me2. The histogram represents the profile of the isochore families, reported as a reference.
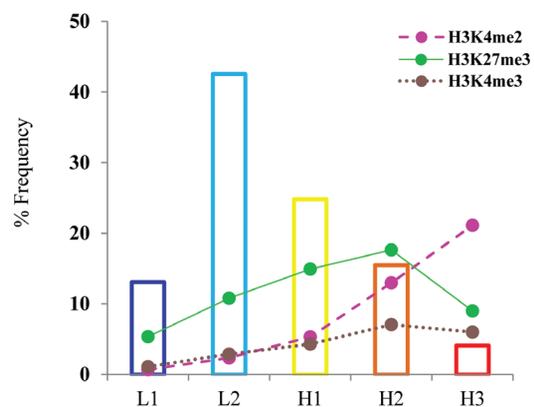
Fig. 5. Plot showing densities of modified histones in the isochore families. Densities (number of genes/Mb of isochore) of the three chromatin marks across isochores in human chromosomes. Brown dotted line: H3K4me3; green full line: H3K27me3; purple dashed line: H3K4me2.
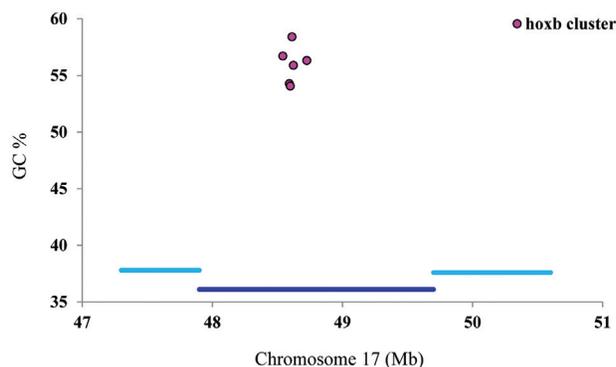
Fig. 6. Genes of Hox B cluster along chromosome 17. Comparison between GC content of *hoxb* genes and of their isochore environment. HOXB cluster is embedded in a L1 isochore that is located between two L2 isochores and has got an average GC level of 36%. The average GC level of *hoxb* genes, instead, is around 55%.

Localization of HOX clusters in human isochores

We obtained the coordinates of the four HOX clusters (associated with H3K27me3 mark 24) from the GENECARDS website (http://www.genecards.org) and determined their position along isochores. We observed that HOXB and HOXD reside in the L2 isochore, whereas HOXA and HOXC belong to the L1 isochore. The four loci are, then, all embedded in GC-poor isochores. In Figure 6, we have reported the example of the HOXB cluster in chromosome 17; it is clear how *hoxb* genes, despite being GC-rich, reside in a GC-poor isochore. We also analyzed the remaining HOX clusters and obtained a similar result (data not shown).

Discussion

Sperm chromatin has attracted the interest of the scientific community for a long time, especially for its peculiar structure and its relevance for studies on infertility, totipotency and embryo development. The prevailing view that the packaging of the paternal genome has a functional meaning has been supported by a number of lines of evidence; for example, abnormalities in protamine and histone patterning are linked to infertility in both mouse and human and to incorrect embryo development.[45–48] Furthermore, it has been reported that histones in sperm chromatin are characterized by a level of plasticity similar to that of embryonic stem cells, thanks to epigenetic marks and posttranslational modifications,[11;49–51] then, they might have a role in transgenerational epigenetic inheritance. Despite being a subject of such interest, our knowledge on the function and localization of determinants of sperm chromatin is still incomplete. In fact, as mentioned in the Introduction, while some studies suggest that histones are bound mostly to promoters and genes,[24,30,33,34] other studies propose that histone retention in sperm is more frequent in repeats and gene poor regions.[12,21,36,52]

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 base composition.[12,24] Our results show that both sequences and a large part of genes associated with mononucleosomes, H3K27me3 and H3K4me3 are located in GC-poor isochores L2 and H1 (~68% of the human genome). Nonetheless, when we consider the density

of genes retaining histones, we observe that it presents an increasing trend from the GC-poor compartment to the GC-rich one, possibly because in the latter the gene density is higher.[39]

For interpretation of the results presented herein, we have to take into account that: i) approximately half of the genes (including genes with high GC%) reside in the GC-poor, gene-poor compartment[39]; ii) in embryo samples, expression begins prevalently from genes located in GC-poor isochores[53]; and, iii) genes constitute only a small percent of the genome,[39] therefore it is expected to find nucleosomes also in nongenic regions. To this regard, the association of nucleosomes with repetitive sequences could have a role in postfertilization processes reviewed in Castillo *et al*,[13] in tridimensional organization of male gamete chromatin and in inheritance of paternal chromatin structure.[54]

Interestingly, regarding point i) above, if we compare the plot showing nucleosome retention with the isochore plot,[12,43] 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,[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.[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.[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.[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.[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.[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.[60]

Perspectives

This work opens new perspectives on the importance of isochore 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.[61] 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

[1] Ward WS. Function of sperm chromatin structural elements in fertilization and development. Mol Hum Reprod 2010;16:30–36. doi:10.1093/molehr/gap080.

[2] Balhorn R. A model for the structure of chromatin in mammalian sperm. J Cell Biol 1982;93:298–305. doi:10.1083/jcb.93.2.298.

[3] Rathke C, Baarends WM, Awe S, Renkawitz-Pohl R. Chromatin dynamics during spermiogenesis. BBA Gene Regul Mech 2014;1839:155–168. doi:10.1016/j.bbagrm.2013.08.004.

[4] Ausió J, González-Romero R, Woodcock CL. Comparative structure of vertebrate sperm chromatin. J Struct Biol 2014;188:142–155. doi:10.1016/j.jsb.2014.09.004.

[5] Saperas N, Chiva M, Pfeiffer DC, Kasinsky HE, Ausió J. Sperm nuclear basic proteins (SNBPs) of agnathans and chondrichthyans: variability and evolution of sperm proteins in fish. J Mol Evol 1997;44(4):422–431. doi:10.1007/pl00006162.

[6] Ausió J, Saperas N, Chiva M. Sperm nuclear basic proteins and sperm chromatin organization in fish. In: Tiersch T, Green C, ed. Cryopreservation in aquatic species 2009, 2nd edition. Baton Rouge, Louisiana, World Aquaculture Society.

[7] Ausió J. The shades of gray of the chromatin fiber: recent literature provides new insights into the structure of chromatin. Bioessays 2015;37(1):46–51. doi:10.1002/bies.201400144.

[8] Oliva R, Dixon GH. Vertebrate protamine genes and the histone to-protamine replacement reaction. Prog Nucleic Acid Res Mol Biol 1991;40:25–94. doi:10.1016/s0079-6603(08)60839-9.

[9] Wu SF, Zhang H, Cairns BR. Genes for embryo development are packaged in blocks of multivalent chromatin in zebrafish sperm. Genome Res 2011;21:578–589. doi:10.1101/gr.113167.110.

[10] Lewis JD, Song Y, de Jong M, Bagha S. Ausió J. A walk though vertebrate and invertebrate protamines. Chromosoma 2003;111:473–482. doi:10.1007/s00412-002-0226-0.

[11] Luense LJ, Wang X, Schon SB, Weller AH, Lin Shiao E, Bryant JM, *et al*. Comprehensive analysis of histone post-translational modifications in mouse and human male germ cells. Epigenetics Chromatin 2016;9:24. doi:10.1186/s13072-016-0072-6.

[12] Samans B, Yang Y, Krebs S, Sarode GV, Blum H, Reichenbach M, *et al*. Uniformity of nucleosome preservation pattern in mammalian sperm and its connection to repetitive DNA elements. Dev Cell 2014;30:23–35. doi:10.1016/j.devcel.2014.05.023.

[13] Castillo J, Estanyol JL, Ballescà JL, Oliva R. Human sperm chromatin epigenetic potential: genomics, proteomics and male infertility. Asian J Androl 2015;17:601–609. doi:10.4103/1008-682x.153302.

[14] Nadel B, de Lara J, Finkernagel SW, Ward WS. Cell-specific organization of the 5S ribosomal RNA gene cluster DNA loop domains in spermatozoa and somatic cells. Biol Reprod 1995;53:1222–1228. doi:10.1095/biolreprod53.5.1222.

[15] Martins RP, Ostermeier GC, Krawetz SA. Nuclear matrix interactions at the human protamine domain: a working model of potentiation. J Biol Chem 2004;279:51862–51868. doi:10.1074/jbc.m409415200.

[16] Hud NV, Downing KH, Balhorn R. A constant radius of curvature model for the organization of DNA in toroidal condensates. Proc Natl Acad Sci U S A 1995;92:3581–3585. doi:10.1073/pnas.92.8.3581.

[17] Gatewood JM, Cook GR, Balhorn R, Bradbury EM, Schmid CW. Sequence-specific packaging of DNA in human sperm chromatin. Science 1987;236:962–964. doi:10.1126/science.3576213.

[18] Gatewood JM, Cook GR, Balhorn R, Schmid CW, Bradbury EM. Isolation of four core histones from human sperm chromatin representing a minor subset of somatic histones. J Biol Chem 1990;265:20662–20666.

[19] Hecht NB. Regulation of 'haploid expressed genes' in male germ cells. J Reprod Fertil 1990;88:679–693. doi:10.1530/jrf.0.0880679.

[20] Banerjee S, Smallwood A, Hultén M. ATP-dependent reorganization of human sperm nuclear chromatin. J Cell Sci 1995;108:755–765.

[21] Pittoggi C, Renzi L, Zaccagnini G, Cimini D, Degrassi F, Giordano R, *et al*. A fraction of mouse sperm chromatin is organized in nucleosomal hypersensitive domains enriched in retroposon DNA. J Cell Sci 1999;112:3537–3548.

[22] Gineitis AA, Zalenskaya IA, Yau PM, Bradbury EM, Zalensky AO. Human sperm telomere-binding complex involves histone H2B and secures telomere membrane attachment. J Cell Biol 2000;151:1591–1598. doi:10.1083/jcb.151.7.1591.

[23] Churikov D, Siino J, Svetlova M, Zhang K, Gineitis A, Morton Bradbury E, *et al*. Novel human testis-specific histone H2b encoded by the interrupted gene on the X chromosome. Genomics 2004;84:745–756. doi:10.1016/j.ygeno.2004.06.001.

[24] Hammoud SS, Nix DA, Zhang H, Purwar J, Carrell DT, Cairns BR. Distinctive chromatin in human sperm packages genes for embryo development. Nature 2009;460:473–478. doi:10.1038/nature08162.

[25] Suzuki T, Minami N, Kono T, Imai H. Comparison of the RNA polymerase I-, II- and III-dependent transcript levels between nuclear transfer and *in vitro* fertilized embryos at the blastocyst stage. J Reprod Dev 2007;53:663–671. doi:10.1262/jrd.19014.

[26] Zheng Z, Jia JL, Bou G, Hu LL, Wang ZD, Shen XH, *et al*. rRNA genes are not fully activated in mouse somatic cell nuclear transfer embryos. J

Biol Chem 2012;287:19949–19960. doi:10.1074/jbc.m112.355099.

[27] van de Werken C, van der Heijden GW, Eleveld C, Teeuwssen M, Albert M, Baarends WM, *et al*. Paternal heterochromatin formation in human embryos is H3K9/HP1 directed and primed by sperm-derived histone modifications. Nat Commun 2014;5:1–15. doi:10.1038/ncomms6868.

[28] Gannon JR, Emery BR, Jenkins TG, Carrell DT. The sperm epigenome: implications for the embryo. Adv Exp Med Biol 2014;791:53–66. doi:10.1007/978-1-4614-7783-9_4.

[29] Nakamura T, Liu YJ, Nakashima H, Umehara H, Inoue K, Matoba S, *et al*. PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. Nature 2012;486:415–419. doi:10.1038/nature11093.

[30] Gardiner-Garden M, Ballesteros M, Gordon M, Tam PP. Histone and protamine-DNA association: conservation of different patterns within the b-globin domain in human sperm. Mol Cell Biol 1998;18:3350–3356. doi:10.1128/mcb.18.6.3350.

[31] Wykes SM, Krawetz SA. The structural organization of sperm chromatin. J Biol Chem 2003;278:29471–29477. doi:10.1074/jbc.m304545200.

[32] Nahkuri S, Taft RJ, Mattick JS. Nucleosomes are preferentially positioned at exons in somatic and sperm cells. Cell Cycle 2009;8:3420–3424. doi:10.4161/cc.8.20.9916.

[33] Arpanahi A, Brinkworth M, Iles D, Krawetz SA, Paradowska A, Platts AE, *et al*. Endonuclease-sensitive regions of human spermatozoal chromatin are highly enriched in promoter and CTCF binding sequences. Genome Res 2009;19:1338–1349. doi:10.1101/gr.094953.109.

[34] Brykczynska U, Hisano M, Erkek S, Ramos L, Oakeley EJ, Roloff TC, *et al*. Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. Nat Struct Mol Biol 2010;17:679–687. doi:10.1038/nsmb.1821.

[35] Vavouri T, Lehner B. Chromatin organization in sperm may be the major functional consequence of base composition variation in the human genome. PLoS genetics 2011;7:e1002036. doi:10.1371/journal.pgen.1002036.

[36] van der Heijden GW, Derijck AA, Ramos L, Giele M, van der Vlag J, de Boer P. Transmission of modified nucleosomes from the mouse male germline to the zygote and subsequent remodeling of paternal chromatin. Dev Biol 2006;298:458–469. doi:10.1016/j.ydbio.2006.06.051.

[37] Royo H, Stadler MB, Peters AHFM. Alternative computational analysis shows no evidence for nucleosome enrichment at repetitive sequences in mammalian spermatozoa. Dev Cell 2016;37:98–104. doi:10.1016/j.devcel.2016.03.010.

[38] Dansranjavin T, Schagdarsurengin U. The rationale of the inevitable, or why is the consideration of repetitive DNA elements indispensable in studies of sperm nucleosomes. Dev Cell 2016;37:13–14. doi:10.1016/j.devcel.2016.03.020.

[39] Bernardi G. Structural and evolutionary genomics. Natural selection in genome evolution. Elsevier, 2004.

[40] Arhondakis S, Auletta F, Bernardi G. Isochores and the regulation of gene expression in the human genome. Genome Biol Evol 2011;3:1080–1089. doi:10.1093/gbe/evr017.

[41] Costantini M, Clay O, Auletta F, Bernardi G. An isochore map of human chromosomes. Genome Res 2006;16:536–541. doi:10.1101/gr.4910606.

[42] Varriale A, Bernardi G. Distribution of DNA methylation, CpGs, and CpG islands in human isochores. Genomics 2009;95:25–28. doi:10.1016/j.ygeno.2009.09.006.

[43] Paces J, Zika R, Paces V, Pavlicek A, Clay O, Bernardi G. Representing GC variation along eukaryotic chromosomes. GENE 2004;333:135–141. doi:10.1016/j.gene.2004.02.041.

[44] Jung YH, Sauria MEG, Lyu X, Cheema MS, Ausiò J, Taylor J, *et al*. Chromatin states in mouse sperm correlatewith embtuyonic and adult regulatory landscapes. Cell Rep 2017;18(6):1366–1382. doi:10.1016/j.celrep.2017.01.034.

[45] Cho C, Willis WD, Goulding EH, Jung-Ha H, Choi YC, Hecht NB, *et al*. Haploinsufficiency of protamine-1 or -2 causes infertility in mice. Nat Genet 2001;28:82–86. doi:10.1038/ng0501-82.

[46] Hammoud SS, Nix DA, Hammoud AO, Gibson M, Cairns BR, Carrell DT. Genome-wide analysis identifies changes in histone retention and epigenetic modifications at developmental and imprinted gene loci in the sperm of infertile men. Hum Reprod 2011;26(9):2558–2569. doi:10.1093/humrep/der192.

[47] Kimura Y, Yanagimachi R. Mouse oocytes injected with testicular spermatozoa or round spermatids can develop into normal offspring. Development 1995;121:2397–2405.

[48] Teperek M, Simeone A, Gaggioli V, Miyamoto K, Allen GE, Erkek S, *et al*. Sperm is epigenetically programmed to regulate gene transcription in embryos. Genome Res 2016;26(8):1034–1046. doi:10.1101/gr.201541.115.

[49] Sassone-Corsi P. Unique chromatin remodeling and transcriptional regulation in spermatogenesis. Science 2002;296:2176–2188. doi:10.1126/science.1070963.

[50] Carrell DT. Epigenetics of the male gamete. Fertil Steril 2012;97:267–274. doi:10.1016/j.fertnstert.2011.12.036.

[51] Castillo J, Amaral A, Azpiazu R, Vavouri T, Estanyol JM, Ballesca JL, *et al*. Genomic and proteomic dissection and characterization of the human sperm chromatin. Mol Hum Reprod 2014;20:1041–1053. doi:10.1093/molehr/gau079.

[52] Carone BR, Hung JH, Hainer SJ, Chou MT, Carone DM, Weng Z. High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. Dev Cell 2014;30(1):11–22. doi:10.1016/j.devcel.2014.05.024.

[53] Frousios K, Iliopoulos CS, Tischler G, Kossida S, Pissis SP, Arhondakis S. Transcriptome map of mouse isochores in embryonic and neonatal cortex. Genomics 2013;101(2):120–124. doi:10.1016/j.ygeno.2012.11.006.

[54] Zalensky A, Zalenskaya I. Organization of chromosomes in spermatozoa: an additional layer of epigenetic information? Biochem Soc Transactions 2007;35:609–611. doi:10.1042/bst0350609.

[55] Barton C, Iliopoulos CS, Pissis SP, Arhondakis S. Transcriptome activity of isochores during preimplantation process in human and mouse. FEBS Lett 2016;590(14):2297–2306. doi:10.1002/1873-3468.12245.

[56] Kel A, Kel-Margoulis O, Babenko V, Wingender E. Recognition of NFATp/AP-1 composite elements within genes induced upon the activation of immune cells. J Mol Biol 1999;288:353–376. doi:10.1006/jmbi.1999.2684.

[57] Hiratani I, Leskovar A, Gilbert DM. Differentiation-induced replication-timing changes are restricted to AT-rich/long interspersed nuclear element (LINE)-rich isochores. Proc Natl Acad Sci USA 2004;101:16861–16866. doi:10.1073/pnas.0406687101.

[58] Kikuta H, Laplante M, Navratilova P, Komisarczuk AZ, Engström PG, Fredman D, *et al*. Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. Genome Res 2007;17:545–555. doi:10.1101/gr.6086307.

[59] Navratilova P, Becker TS. Genomic regulatory blocks in vertebrates and implications in human disease. Brief Funct Genomic Proteomic 2009;8:333–342. doi:10.1093/bfgp/elp019.

[60] Siklenka K, Erkek S, Godmann M, Lambrot R, McGrawS, Lafleur C, Cohen T, Xia J, Suderman M, Hallett M, *et al*. Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. Science 2015;350(6261):aab2006. doi:10.1126/science.aab2006.

[61] Wu S, Zhang H, Cairns BR. Genes for embryo development are packaged in blocks of multivalent chromatin in zebrafish sperm. Genome Res 21:578–589. doi:10.1101/gr.113167.110.