Genomics xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Distribution of DNA methylation, CpGs, and CpG islands in human isochores

Annalisa Varriale, Giorgio Bernardi *

Laboratory of Molecular Evolution; Stazione Zoologica Anton Dohrn, 80121 Naples, Italy

ARTICLE INFO

Article history:
Received 2 July 2009
Accepted 25 September 2009
Available online xxxx

Keywords:
CpG
DNA methylation
Human genome
Nucleotide composition

ABSTRACT

DNA methylation is a major epigenetic modification of the genome that affects basic biological functions, such as gene expression and cell development. We used the human genome sequences and the DNA methylation data that are available in order to establish a map of the levels of GC and methylation in isochores. We also looked for the correlations that hold between GC levels and the distribution of the (1) dinucleotide CpG, (2) ratio 5mC/CpG, and (3) CpG islands. Our results show that methylation levels, CpG frequencies, and the density of CpG islands are positively correlated with the GC level of isochores. In contrast, the correlation between the 5mC/CpG ratio and GC is a negative one because the increase in methylation lags behind that of CpG, to reach a plateau in the GC-richest, gene-richest isochore families H2 and H3. In conclusion, there are more CpG targets that remain unmethylated in the GC-richest, gene-richest isochores in comparison with the other isochores. This conclusion supports the idea that the widespread methylation under consideration here has a general inhibitory effect on gene expression.

Introduction

For vertebrates, DNA methylation is a major epigenetic modification, which has a role at both the genic and the genomic levels (see references [1–3] for recent reviews). At the gene level, one of the functions of methylation is to regulate gene expression, and many such cases have been reported in the literature (see, for example, a classical review by Razin and Cedar [4], as well as by Robertson [5]). In contrast, even as the number of techniques for genome-wide analysis of DNA methylation has increased [6], results on methylation at the genome level are still limited. Moreover, because methylation takes place almost exclusively at the cytosines located in CpG nucleotides, the study of the distribution of DNA methylation cannot be separated from that of CpG. For vertebrate genomes, however, existing DNA methylation data reveal that this modification is not randomly distributed (see below).

In the mid-1980s, we showed that CpG is linearly and positively correlated with GC levels of vertebrate genes ([7] see also Bernardi and Bernardi [8]), a point confirmed by further detailed investigations [9]. We then showed that CpG islands (sequences of 0.5–2 kb in size that are rich in GC and unmethylated CpG doublets) were scarce or absent in cold-blooded vertebrates [10,11], whereas in warm-blooded vertebrates, their densities almost paralleled the increase of gene densities in isochores of increasing GC [9]. Methylation linearly increased with genome GC in the case of all vertebrates tested, but methylation levels of warm-blooded vertebrates were systematically lower than those of cold-blooded vertebrates [12–14]. This correlation also held intragenomically as investigated in *Xenopus*, chicken, mouse, and human [15]. Finally, a negative correlation was found between 5mC/CpG and GC because the increase in CpG was steeper than that of 5mC [9].

More recently, the complete sequences of the human genome allowed establishing the methylation level [16] and an isochore map [17] along human chromosomes. These two resources gave us the possibility of providing the complete pattern of methylation along the isochores of the human genome and of defining the correlations that hold between isochores and methylation, CpGs, and CpG islands.

Materials and methods

Correspondence between methylation in the clones of Weber et al. and the isochore bands of Costantini et al.

We used the methylation data obtained by immunoprecipitating methylated DNA (MeDIP) in clones of about 80 kb that covered all the human genome [16]. For each clone, the authors provided the position on the chromosomes (UCSC release hg15; April 2003). We also collected the coordinates of the isochore borders as described in Costantini et al. [17]. These authors mapped about 3200 isochores on the fully sequenced human genome (UCSC release hg17; May 2004), which represented a complete coverage of the chromosomes (neglecting the remaining gaps). In order to make the exact match of methylation data with the isochores, we had to find the relative position of each clone within the isochore map for all chromosomes. Given that between the two releases there are some differences due to gaps, the coordinates of the clones often did not coincide with the right isochore; thus, we had to shift them manually to find the right

* Corresponding author. Fax: +39 081 2455807.
E-mail address: bernardi@szn.it (G. Bernardi).

0888-7543/$ – see front matter © 2009 Published by Elsevier Inc.
doi:10.1016/j.ygeno.2009.09.006

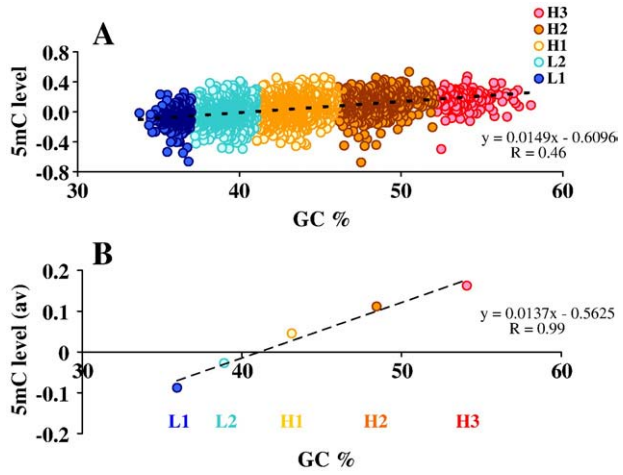

Fig. 1. Correlation between 5mC and GC levels in the isochores of the human genome. We assembled the data for all chromosomes, except Y. (A) The coordinates of each point are the GC level of the isochores [17] and the average 5mC level for each of the group of clones analyzed by [16] as present in each isochore. The 5mC level was represented as log ratio of methylated to input signal for each clone spotted on microarray. (B) Correlation between 5mC and GC levels in isochore families. The colors used for isochore families are indicated. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

match. At the end of this process, we calculated the average GC and methylation levels for all isochores in all chromosomes, except for the Y chromosome, for which methylation data were not available.

Evaluation of the CpG frequency along the isochores

The frequency of CpG was calculated with a PERL script implemented in our laboratory. This script counted the number of CpG doublets in all isochores. For each isochore, we associated the corresponding methylation level (see above) from Weber et al. [16] and evaluated the densities of both.

CpG island frequency along the isochores

In order to evaluate the contribution of CpG islands to the ratio of CpG frequency over methylation, we downloaded the file of the

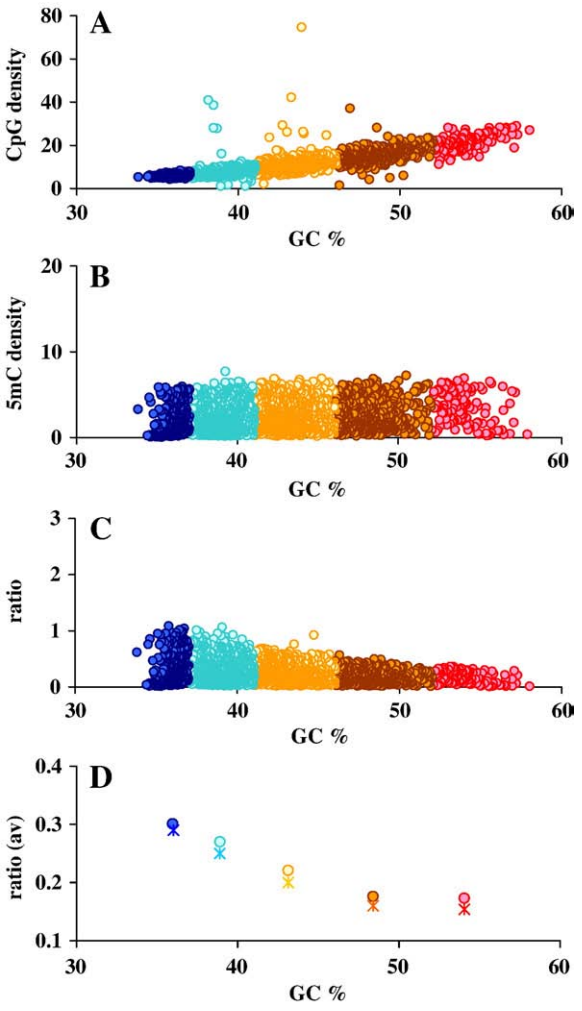

Fig. 3. Distribution of CpG, 5mC, and 5mC/CpG density ratio. (A) Distribution of the CpG density. (B) Distribution of 5mC density, (C) distribution of the 5mC/CpG ratio. (D) Circles and asterisks represent the average ratio for isochore families excluding and including the contribution of CpG islands, respectively.

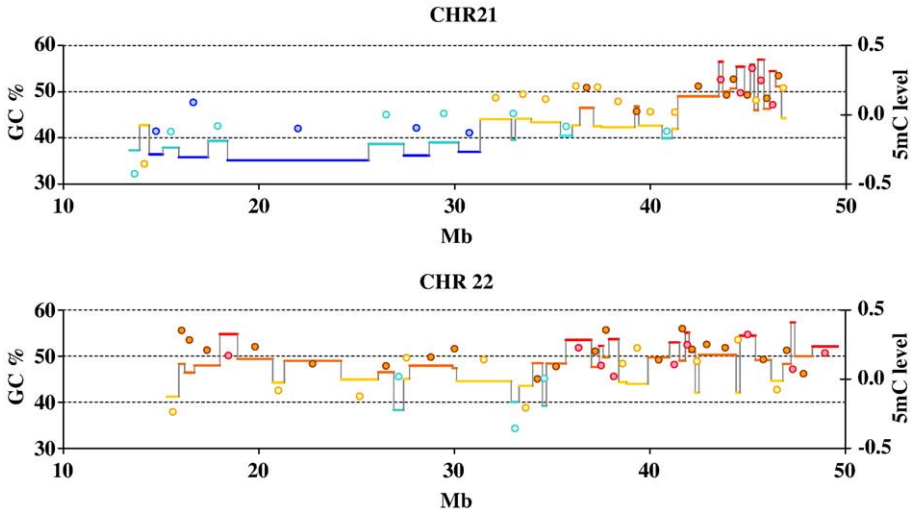

Fig. 2. Display of methylation level along the isochores of chromosomes 21 and 22. The horizontal bars indicate the isochores as obtained by [17], whereas the circles indicate the 5mC level (see legend of Fig. 1). In both cases, the color code is that used in Fig. 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Please cite this article as: A. Varriale, G. Bernardi, Distribution of DNA methylation, CpGs, and CpG islands in human isochores, Genomics (2009), doi:10.1016/j.ygeno.2009.09.006

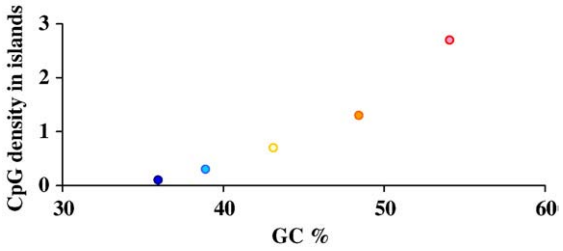

Fig. 4. Plot of the density of CpGs in CpG islands for the isochore families. The color code follows that used in the previous figures. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

coordinates of the CpG islands from the release hg17 of the UCSC Website. We then filtered these "putative" islands with the criteria of Takai and Jones [22]. We then assigned each island to the corresponding isochore and used the PERL script to calculate the new CpG frequency. Once downloaded and filtered, we assessed the frequency of the CpG islands in human isochores by assigning each island to the corresponding isochore through their coordinates.

Choice of the criteria for establishing the CpG islands

Various authors have proposed different definitions for CpG islands based on sequence criteria for the *in silico* prediction of islands, even if a direct experimental evaluation of methylation status is required in order to get the best definition. The original criteria defined them as regions of at least 200 bp with a GC level of at least 50% and a ratio of CpG observed to expected (o/e) of at least 0.6 [23]. Recently, such criteria were refined by Takai and Jones [22], such that most Alu repeats were excluded. We decided to use their criteria, as it was demonstrated that their parameters allow many false positives to be discarded. Another possibility would have been to use the CpG islands found by other authors [24], who constructed CpG island maps also based on epigenomic features; however, the authors themselves stressed that their method still needs further research.

Results

Correlation between methylation and GC levels of isochores

Fig. 1 summarizes our results on the correlation between methylation and GC level in the isochores from human chromosomes. Isochores were divided into GC-poor, gene-poor (L1, L2, and H1), and GC-rich, gene-rich (H2 and H3) families. Fig. 1A shows the results for isochores, with each point indicating the average methylation and GC level for each isochore, whereas Fig. 1B shows the average levels for both methylation and GC for the isochore families.

As an example, we reported the linear distribution of methylation levels along chromosomes 21 and 22, as partitioned in isochores (Fig. 2). It is clear already from these two chromosomes that the methylation level is not strictly proportional to GC level when judged at the isochore level. Such variations are very frequent as it can be seen in the full display of methylation results for all chromosomes (Supplemental Figure S1).

Correlation between the 5mC/CpG ratio and the GC levels of isochores

Since CpG islands are known to be unmethylated in most cases, we calculated the CpG densities and the ratio of CpG and 5mC densities after the exclusion of the CpG islands from each isochore band. We reported the results of this analysis for both 5mC and CpG density in isochores, together with their ratio on a GC level scale in Fig. 3. More in particular, the densities of CpG and 5mC are shown in Figs. 3A and B, respectively, whereas Figs. 3C and D displays the 5mC/CpG density ratio. In particular, panel C shows the points for each isochore, and panel D, the average values for the five isochore families. Interestingly, if we compare this result with that obtained with all CpGs, including those of the islands, we notice that the two results are very similar (panel D). The explanation can be found in the plot of the contribution of the CpGs in the CpG islands for each isochore family (Fig. 4). Such contribution increases linearly with the GC level, yet it does not change the trend of CpG density in isochores. In both cases, in fact, the CpG density, unlike the 5mC density, increases with the increasing of GC level. The ratio of the two 5mC/CpG densities, on the contrary, declines with increasing GC of isochores because the increase in methylation with isochore GC lags behind the increase of CpG until a 5mC plateau is reached.

CpG island density in human isochores

When we analyzed the number and density of CpG islands in isochores, we observed that both positively correlate with the GC level, as is in the case for genes (Fig. 5). It is well known, in fact, that CpG islands are often associated with promoters and that genes reach a maximum concentration in H3 isochores.

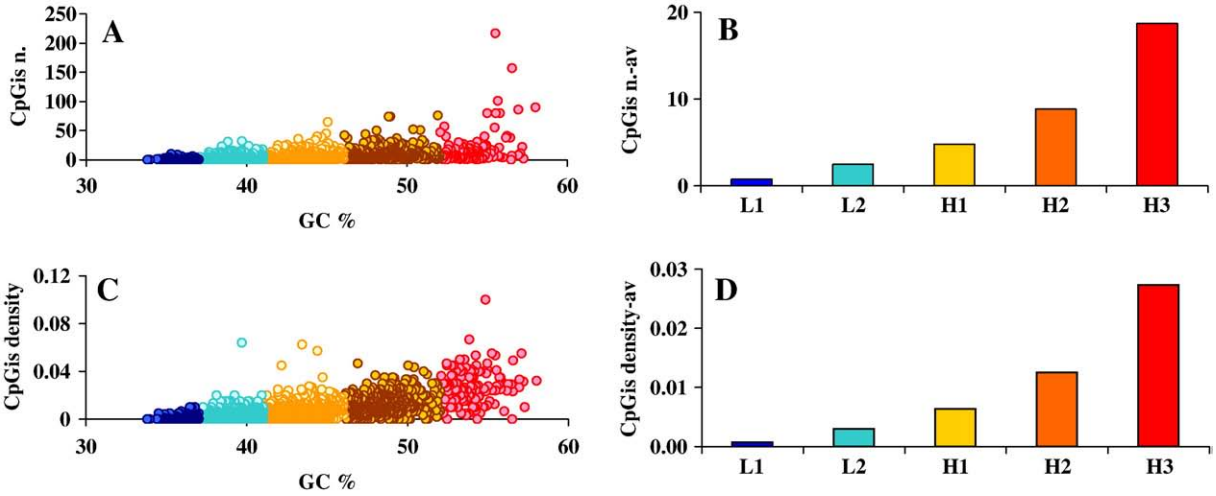

Fig. 5. Frequency and density of CpG islands in the isochores. The frequency is the number of islands for each isochore, whereas the density is calculated considering the size of isochores. (A) Distribution of CpG islands in isochores. (B) Average frequency of CpG islands in isochore families. (C) Density of CpG islands in isochores. (D) Average density of CpG islands in isochore families.

Please cite this article as: A. Varriale, G. Bernardi, Distribution of DNA methylation, CpGs, and CpG islands in human isochores, Genomics (2009), doi:10.1016/j.ygeno.2009.09.006

Discussion

We have reported here an analysis carried out for the first time on the distribution of CpG doublets, 5mC, 5mC/CpG ratios, and CpG islands on a fully sequenced mammalian genome (i.e., the human genome) in which methylation was precisely assessed.

These results lead to three major conclusions. First, they confirm our previous investigations that were done on compositional DNA fractions, genes, and the limited number of available sequences (see Introduction). Indeed, using the new data, we showed that (i) a positive linear correlation holds between the frequency of CpG and 5mC and the GC levels of isochores; (ii) a negative correlation links the 5mC/CpG ratio and the GC levels of isochores, with a plateau in the GC-richest, gene-richest isochore families H2 and H3; and (iii) a positive linear correlation exists between the frequency of CpG islands and the GC levels of isochores.

Obviously, while all these points were already established qualitatively, with the exception of the plateau level of 5mC/CpG ratio, the present results quantify them definitively. Of particular interest are the findings that (i) the correlation between CpG and isochore GC level does not change significantly whether unmethylated CpG islands are or not excluded; (ii) CpG islands, namely, sequences that have a regulatory role, increase in parallel with genes as GC levels of isochores increase; and (iii) the 5mC/CpG ratio reaches a plateau because the 5mC increase with increasing isochore GC lags behind the increase of CpG.

The second conclusion is that the latter point may be understood by thinking that the overall DNA methylation, which is due to DNMT1 methylase [18], has a general inhibitory role on gene expression and therefore its frequency is decreased in the gene-richest isochores. This second conclusion goes along with our observation that global DNA methylation level in polar fishes is higher in comparison with temperate and tropical species [13]. Indeed, DNA methylation could be used to block large genomic regions that remain unexpressed since the activity of the genes located in them is not needed. This idea is supported by the observation that Antarctic fishes lose the expression of some genic regions [19,20].

The third conclusion is that while general rules about the distribution of methylation clearly exist, regional variations can also be observed. This is understandable because of local variations in the density of genes and interspersed repeats (Alu sequences are fully methylated). This stresses the interest of the maps presented here. Indeed, any region of particular concern can be enlarged, observed in detail, and allow linking methylation level, gene position, and regional gene expression.

A final comment is that, so far, isochores were associated with basic properties of the genome, such as gene density, chromatin structure, replication timing, and recombination [1], to name only the most important of them, so justifying calling isochores "a fundamental level of genome organization" [21]. The present work indicates that isochores belonging to different families can also be distinguished on the basis of differences in 5mC, CpGs, 5mC/CpG and CpG island densities.

Conclusions

This study represents an analysis of the distribution of methylation, CpGs, and CpG islands in the human isochores. It shows highly significant correlations between the three parameters and the GC level of isochores, as well as a negative correlation between the ratio 5mC/CpG and GC, meaning that in GC-rich, gene rich-regions many targets remain unmethylated. Finally, this work provides maps of methylation along the chromosomes that can be used to find out the methylation level of each isochore and correlate it with gene expression.

Acknowledgments

We wish to thank Dirk Schübeler for his help with the file of methylation data, Maria Costantini for the original data of the isochore maps of chromosomes, and Fabio Auletta and Giuseppe Torelli for the PERL scripts.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2009.09.006.

References

[1] G. Bernardi, Structural and Evolutionary Genomics. Natural Selection in Genome Evolution, : Elsevier, Amsterdam, 2005 (reprinted from 2004).
[2] B.E. Bernstein, A. Meissner, E.S. Lander, The mammalian epigenome, Cell 128 (2007) 669–681.
[3] M. Weber, D. Schübeler, Genomic patterns of DNA methylation: targets and function of an epigenetic mark, Curr. Opin. Cell Biol. 19 (2007) 273–280.
[4] A. Razin, H. Cedar, DNA methylation and gene expression, Microbiol. Mol. Biol. Rev. 55 (1991) 451–458.
[5] K.D. Robertson, DNA methylation and human disease, Nat. Rev., Genet. 6 (2005) 597–610.
[6] D. Zilberman, S. Henikoff, Genome-wide analysis of DNA methylation patterns, Development 134 (2007) 3959–3965.
[7] G. Bernardi, B. Olofsson, J. Filipski, M. Zerial, J. Salinas, G. Cuny, M. Meunier-Rotival, F. Rodier, The mosaic genome of warm-blooded vertebrates, Science 228 (1985) 953–958.
[8] G. Bernardi, G. Bernardi, Compositional constraints and genome evolution, J. Mol. Evol. 24 (1986) 1–11.
[9] K. Jabbari, G. Bernardi, CpG doublets, CpG islands and Alu repeats in long human DNA sequences from different isochore families, Gene 224 (1998) 123–128.
[10] B. Aïssani, G. Bernardi, CpG islands: features and distribution in the genome of vertebrates, Gene 106 (1991) 173–183.
[11] B. Aïssani, G. Bernardi, CpG islands, genes and isochores in the genome of vertebrates, Gene 106 (1991) 185–195.
[12] K. Jabbari, S. Cacciò, J.P. Pais de Barros, J. Desgres, G. Bernardi, Evolutionary changes in CpG and methylation levels in the genome of vertebrates, Gene 205 (1997) 109–118.
[13] A. Varriale, G. Bernardi, DNA methylation and body temperature in fishes, Gene 385 (2006) 111–121.
[14] A. Varriale, G. Bernardi, DNA methylation in reptiles, Gene 385 (2006) 122–127.
[15] S. Cacciò, K. Jabbari, G. Matassi, F. Guermonprez, J. Desgrès, G. Bernardi, Methylation patterns in the isochores of vertebrate genomes, Gene 205 (1997) 119–124.
[16] M. Weber, J.J. Davies, D. Wittig, E.J. Oakeley, M. Haase, W.L. Lam, D. Schübeler, Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells, Nat. Genet. 37 (2005) 853–861.
[17] M. Costantini, O. Clay, F. Auletta, G. Bernardi, An isochore map of human chromosomes, Genome Res. 16 (2006) 536–541.
[18] T.H. Bestor, The DNA methyltransferases of mammals, Hum. Mol. Genet. 9 (2000) 2395–2402.
[19] G.E. Hofmann, B.A. Buckley, S. Airaksinen, J.E. Keen, G.N. Somero, Heat-shock protein expression is absent in the antarctic fish *Trematomus bernacchii* (family *Nototheniidae*), J. Exp. Biol. 203 (2000) 2331–2339.
[20] G.N. Somero, Linking biogeography to physiology: evolutionary and acclimatory adjustments of thermal limits, Front. Zool. 2 (2005) 1–9.
[21] A. Eyre-Walker, L. Hurst, The evolution of isochores, Nat. Rev., Genet. 2 (2001) 549–555.
[22] D. Takai, P.A. Jones, Comprehensive analysis of CpG islands in human chromosomes 21 and 22, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 3740–3745.
[23] M. Gardiner-Garden, M. Frommer, CpG islands in vertebrate genomes, J. Mol. Biol. 196 (1987) 261–282.
[24] C. Bock, J. Walter, M. Paulsen, T. Lengauer, CpG island mapping by epigenome prediction, PLOS Comput. Biol. 3 (2007) e110.

Please cite this article as: A. Varriale, G. Bernardi, Distribution of DNA methylation, CpGs, and CpG islands in human isochores, Genomics (2009), doi:10.1016/j.ygeno.2009.09.006