Genomics 101 (2013) 120–124

Contents lists available at SciVerse ScienceDirect

Genomics

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

Transcriptome map of mouse isochores in embryonic and neonatal cortex

Kimon Frousios [a], Costas S. Iliopoulos [a,b], German Tischler [c,1], Sophia Kossida [d], Solon P. Pissis [e,f], Stilianos Arhondakis [d,*]

[a] Department of Informatics, King's College London, The Strand, London WC2R 2LS, UK
[b] School of Mathematics and Statistics, University of Western Australia, 35 Stirling Highway, Crawley, Perth WA 6009, Australia
[c] Lehrstuhl für Informatik 2, Universität Würzburg, Am Hubland, 97074 Würzburg, Germany
[d] Biomedical Research Foundation of the Academy of Athens, 4 Soranou Ephessiou, Athens 115 27, Greece
[e] Florida Museum of Natural History, University of Florida, 1659 Museum Road, Gainesville, FL 32611, USA
[f] Heidelberg Institute for Theoretical Studies, 35 Schloss-Wolfsbrunnenweg, Heidelberg D-69118, Germany

ARTICLE INFO

Article history:
Received 4 September 2012
Accepted 15 November 2012
Available online 27 November 2012

Keywords:
Base composition
Isochores
GC level
RNA-seq
Expression
Developmental process

ABSTRACT

Several studies on adult tissues agree on the presence of a positive effect of the genomic and genic base composition on mammalian gene expression. Recent literature supports the idea that during developmental processes GC-poor genomic regions are preferentially implicated. We investigate the relationship between the compositional properties of the isochores and of the genes with their respective expression activity during developmental processes. Using RNA-seq data from two distinct developmental stages of the mouse cortex, embryonic day 18 (E18) and postnatal day 7 (P7), we established for the first time a developmental-related transcriptome map of the mouse isochores. Additionally, for each stage we estimated the correlation between isochores' GC level and their expression activity, and the genes' expression patterns for each isochore family. Our analyses add evidence supporting the idea that during development GC-poor isochores are preferentially implicated, and confirm the positive effect of genes' GC level on their expression activity.

1. Introduction

The genomes of vertebrates are mosaics of isochores, long regions (from 0.2 Mb up to several Mb) that are fairly homogeneous in base composition, belonging to a small group of families characterized by different GC levels (molar ratio of guanine and cytosine over the total number of bases of the area; [1–4]). The mouse genome, similarly to the human genome, is characterized by five isochore families (L1, L2, H1, H2 and H3; in order of increasing GC level). The difference though is the under-representation of the L1 family, and the nearly absence of the H3 family in the mouse genome [5]. Moreover, the GC-richest families, representing approximately 15% of the genome, contain approximately 50% of the protein-coding genes, and are associated with several structural and functional properties opposite to those reported for the GC-poorer ones [2]. One such property is gene expression. Until now, most studies [6–13], using expression data from sequencing-based (ESTs, SAGE, MPSS) and/or hybridization-based techniques (microarrays, cDNA arrays), agree on a positive effect of genes' GC3 levels on mammalian gene expression in the adult tissues. In the same direction, although through a different approach, two studies [14,15] suggested that, aside from the GC3 level of the coding sequences, the genomic compositional context in which a gene is embedded affects its expression. Additionally, at the level of genomic regions, clusters of highly and weakly expressed genes have been detected [16,17]. The former were named RIDGEs (regions of increased gene expression), and were located in the gene-dense, GC-rich regions, while the latter ones, anti-RIDGEs, and were identified in gene-poor and low-GC regions. Finally, a recently published study [18], using publicly available RNA-seq data from three different adult tissues of mouse, confirmed the existence of a higher expression activity of the GC-richer isochores and genes compared to that of the low-GC ones.

On the other side, there have been an increasing number of studies supporting the idea that low-GC regions of the genome are active during development, and switch off after chromatin compaction at the end of it. Specifically, it has been shown that, during mouse brain development, most expression changes occur in the genes that are localized in GC-poor, LINE-rich, regions [19]. Similarly, Ren et al. [20] reported that the genes expressed in early developmental stages, compared to the genes expressed in later developmental stages, have a preference toward AT-ending codons (low GC). As GC3 of genes is correlated with the GC level of the isochores [21], those genes would be also typically embedded in low-GC isochores. Moreover, Kikuta et al. [22] and Navratilova and Becker [23] provided evidence of existing conserved systems across vertebrate genomes, characterized by highly conserved non-coding regulatory elements and their target developmental genes, shown elsewhere to be embedded in gene-poor/low GC genomic regions [24].

* Corresponding author.
E-mail address: sarhondakis@bioacademy.gr (S. Arhondakis).
[1] Current address: The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK.

0888-7543/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ygeno.2012.11.006

Motivated by the above studies, we decided to investigate how the base composition of the isochores and of the genes is related to their corresponding expression activity, during developmental processes. Our strategy involved the use of publicly available RNA-seq data from two distinct developmental stages, embryonic day 18 (E18) and postnatal day 7 (P7; [25]), in order to establish, for each of these developmental stages: i) the transcriptome map of the mouse isochores, ii) the relation between isochores' compositional properties and their expression activity, and iii) the relation between the expression activity of the protein coding genes and their location in the isochore families.

2. Results

2.1. Transcriptome map of mouse isochores in the two developmental stages

As in our previous study on adult tissues [18], we first established the isochores' expression profiles (see Eq. (1) in the Material and methods section) for the two developmental stages along the isochores of the whole genome (see Supplementary Fig. 1). For each of these two stages, we can observe a rough agreement of the expression activity of the isochores with their GC levels. One example can be clearly seen on Chromosome 1 (Fig. 1). In order to quantify this relation, we also examined the correlations between the overall expression activity within each isochore (see Eq. (1) in the Material and methods section) and that isochore's respective GC level, and found it to be positive and statistically significant (correlation coefficient in embryo is 0.507 and in postnatal 0.536; see Supplementary Fig. 2). Interestingly, these positive correlations were found to be weaker than those reported in our previous work for adult brain ($R = 0.72$; [18]).

After the overall transcriptome map was established, we decided to focus on the correlation between isochores' GC level and their expression activity. In order to quantify this correlation, it is necessary to eliminate within each isochore the effects of the gene density, known to be higher in the GC-richer isochores of vertebrates. To achieve this, the normalized per developmental stage number of aligned reads within each isochore, was divided by the respective gene density of the isochore, and the $\log_2$ was calculated (see Eq. (2) in the Material and methods section). This approach led to the elimination of 335 isochores with no CDS (enriched in low-GC isochores, L1 and L2; see Supplementary Fig. 3), reducing our initial dataset to 1984 isochores with at least one CDS. We then measured the correlation between the expression level of the isochores (after gene density effects were eliminated) and their respective GC levels, and found no correlation in either the embryonic or postnatal stage (Fig. 2). This is in contrast to the positive correlations we reported in our previous study for the adult brain ($R = 0.22$; [18]).

Summarizing, in this section we report: i) the first transcriptome map of the mouse isochores for two developmental stages (see Fig. 1 and Supplementary Fig. 1) and ii) a lack of correlation between the GC level of the isochores and their expression activity (see Fig. 2).

2.2. Relation between genes' expression activity and their localization in the isochore families

In this section, we first estimated the expression level of each gene for each developmental stage (see Eq. (3) in the Material and methods section). Subsequently, genes were partitioned in the five isochore families, and, for each family, the average genic expression was calculated. For both developmental stages, we found an increase in the average genic expression level as the GC level of the isochore increases (Fig. 3). In order to reduce possible noise from misalignments,

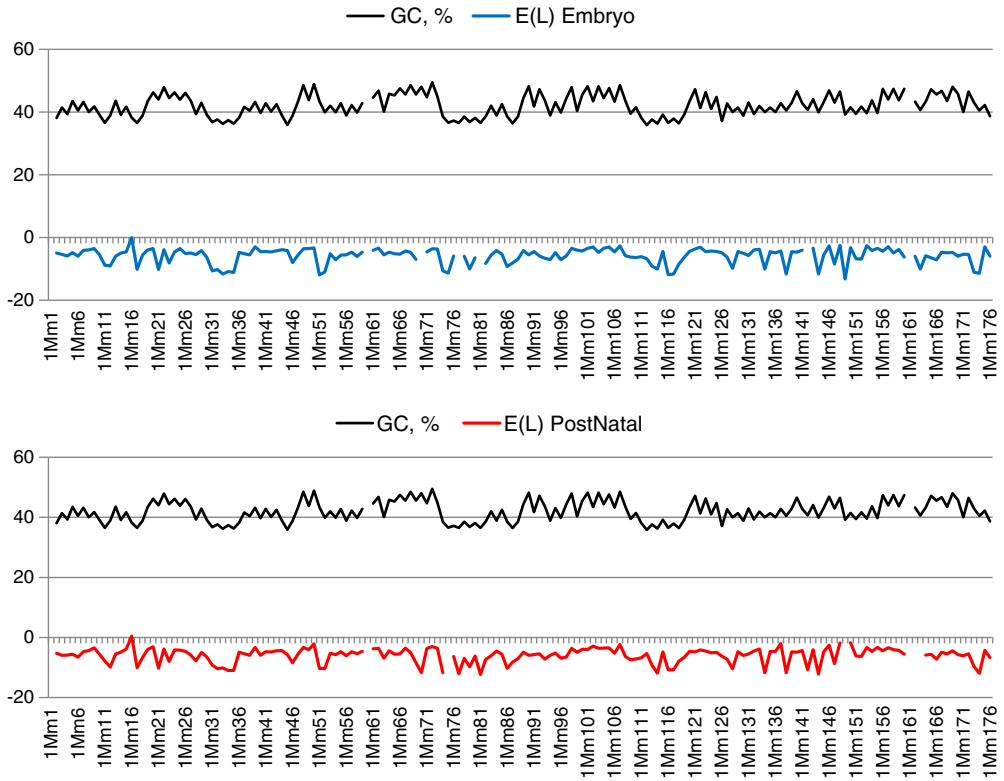

Fig. 1. Expression profiles of the isochores on Chromosome 1 for the two developmental stages. The Y axis measures the isochores' GC levels (positive values; black line) and their respective expression levels (E_L; Eq. (1); negative values) for the embryonic (upper panel; blue line) and postnatal stages (lower panel; red lines). High expression corresponds to peaks in the lines.

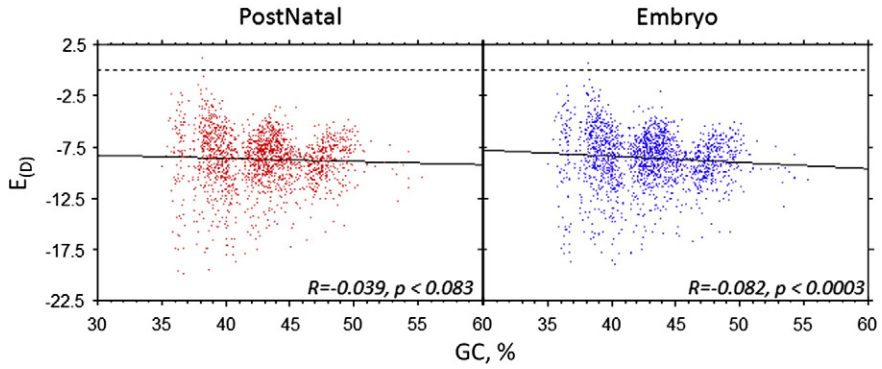

Fig. 2. Correlation between the expression level (normalized by the gene density E_D; Eq. (2)) of each isochore and the respective GC level (blue plot for embryo and red plot for postnatal).

we repeated the above analysis, after eliminating any coding sequences with less than 10 aligned reads. This led to a reduction of the expressed coding sequences to just half of their initial numbers (P7 from 11,838 to 5358 CDSs; E18 from 11,096 to 5088 CDSs). Despite the application of this expression threshold, we were still able to find an increase in average genic expression level as the GC of the isochores increases (see Supplementary Fig. 4). For each stage, we also examined the isochoric distribution of the coding sequences with 0, <10, and ≥10 aligned reads. We found that the coding sequences with <10 and ≥10 aligned reads did not have any significant differences in their distributions (see Supplementary Fig. 5), thus excluding the possibility of GC biases between the number of aligned reads and the base composition of the coding sequences; an effect that could otherwise be responsible for the increased expression of the GC-rich genes. We must note, that our initial datasets of expressed genes in E18 (11,096) and P7 (11,838) stages are in good agreement with the number of expressed genes reported by Han et al. ([25]; 13,463 in E18 and 14,243 in P7), and any differences are due to the use of different databases.

Finally, for each stage, we investigated the compositional preferences (in terms of isochore families) of the expressed genes (≥10 aligned reads), after they were divided into five categories: i) genes with higher expression in postnatal (2420 CDSs), ii) genes exclusively expressed in postnatal stage (1041 CDSs), iii) genes with higher expression in embryonic stage (1898 CDSs), iv) genes exclusively expressed in embryonic stage (771 CDSs), and v) genes not detected to be expressed in any stage (16,001 CDSs). We observed that almost all gene categories have similar distribution in the isochore families, except for the genes expressed exclusively in embryonic stage and those non-detected to be expressed at all, both enriched in genes located in L2 and H1 isochores, instead of H2 isochores (see Supplementary Fig. 6). Again, there seems to be an agreement between the number of genes we detected as exclusively expressed in either P7 (1041 CDSs) or E18 (771) with those identified by Han et al. ([25]; 1296 in P7 and 660 in E18; any differences are due to the use of different databases).

Summarizing, in this section, we show that GC-rich genes have a higher expression compared to the GC-poor genes (Fig. 3 and Supplementary Fig. 4), and that the genes exclusively expressed in the embryonic stage, compared to those expressed either in both stages or exclusively in the postnatal stage (later stage), are preferentially located in GC-poorer isochores (see Supplementary Fig. 6).

3. Discussion

The application of RNA-seq data provides a rapid and cost-effective way to obtain large amounts of transcriptome data for genome-wide investigations. Very recently, using RNA-seq data we were able to establish the first transcriptome map of the mouse isochores from three adult tissues, and to confirm the positive effects of the isochores and genes compositional properties on their corresponding expression activity [18], supporting the appearance of the isochores as a result of evolutionary adaptation process [2,3]. As we anticipated (see Introduction section) there is increasing evidence

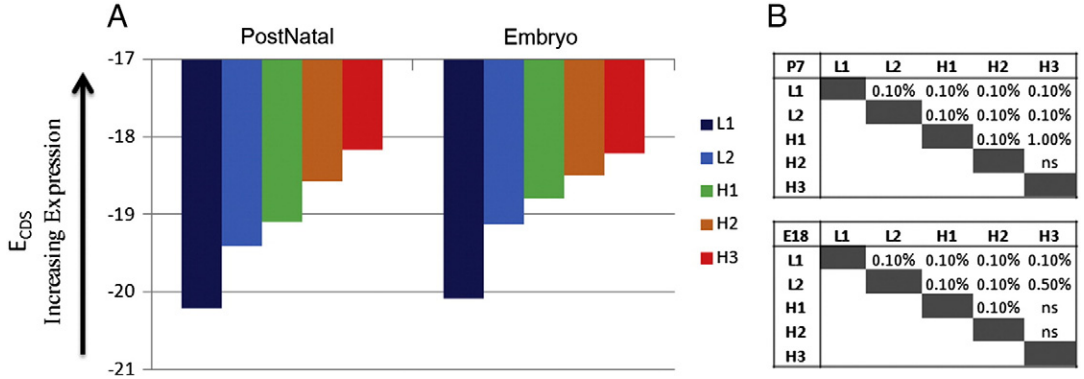

Fig. 3. Panel A: Average genic expression levels after the genes have been binned in the five isochore families. Larger negative values (tall colored bars) indicate low expression, and small negative values (short colored bars) indicate high expression. Panel B: Within each developmental stage the p-values of the average genic expression differences between the isochore families (Cochran test, non-parametric). The upper table is for the postnatal stage (P7) and the lower for the embryonic stage (E18).

supporting the idea that low-GC genomic regions are involved during developmental processes. In the current study, using publicly available RNA-seq data from two distinct developmental stages, we reported the first development-related transcriptome map of the mouse isochores (Fig. 1 and Supplementary Fig. 1), and we showed an absence of correlation between the isochores' expression activity and their GC levels (Fig. 2). This absence of a correlation, is in contrast with what we have previously published for the adult brain in mouse [18], and it can be interpreted as an increased expression activity of the low-GC isochores; reflecting an implication of the low-GC non-coding DNA for an orderly chromatin condensation and a chromatin-mediated activation of the development-specific genes, before its compaction at the end of this process. Finally, the findings at the gene level (Fig. 3 and Supplementary Fig. 4) confirm that during development, as shown for the adult brain (see [18]), the genes located in GC richer isochores have a higher average expression compared to the genes located in GC poorer isochores. Interestingly, we also observed an increasing trend in the GC level of the expressed genes as development advances. Indeed, the genes exclusively expressed in the embryonic stage, when compared to those expressed either in both stages or only in the postnatal stage, were found to be GC-poorer (see Supplementary Fig. 6). This finding, although not possible to ascertain due to the low number of developmental stages, is in line with the observation of Ren et al. [20], who reported that the genes expressed in the early developmental stages, compared to the genes expressed in the later developmental stages, have more AT-ending codons (lower GC).

The functional role of genes' compositional properties on gene expression activity, in different cellular contexts (see Introduction section) and biological processes, i.e., aging [26], and that of the isochores' on the regulation mechanisms, represents a relevant subject of study. Indeed, a recent work [27], reported differences in the frequency of the short sequences (di- and tri-nucleotides) between the isochore families, accounting for their different properties, i.e., codon frequencies, the distribution of DNA methylation, CpG doublets, and CpG islands [28], as well as the chromatin structure [29–31]. In a similar context, it has been suggested that the genes located in different isochore families are subject to different regulation mechanisms, acting at the level of chromatin structure (nucleosome positioning) or of their regulatory regions [32]. Other studies, promote an existing relation between the compositional structure of promoters and genes with their biological functions. Indeed, Carninci et al. [33], have shown that mammalian promoters are separated into two distinct classes, the TATA-box-enriched class and the GC-rich class, each class used differently in the different biological contexts and/or gene families. In the same direction, another investigation [34], reports an existing functional preference of the GC-rich and the GC-poor promoters; the former preferentially involved in binding and protein transport activities, and the latter ones in environmental defense responses. Moreover, for the immune-cell specific promoters it has been shown to be GC poor [35]. Finally, a similar relation between base composition and biological function is also reported at the level of genes. Sandelin et al. [24] reported that genes located in the gene desert (low-GC regions) are enriched in specific gene ontology terms, among them, development. Another work [36], revealed that different functional classes include genes characterized by different compositional properties, i.e., genes involved in cellular metabolism have significantly higher GC3 levels to the genes involved in information storage and processing.

Summarizing, the aforementioned studies, including the present one, support a functional implication of the isochores and genes' compositional properties in different biological contexts. Specifically, our results add new insights on the role of the isochores during developmental process, although more data are necessary in order to fully determine this existing relationship, and re-confirm a positive effect of genes' GC level on mammalian gene expression. Concluding, we propose that a compositional exploration of the expression patterns during biological processes may increase our understanding on how the compositional properties of coding and non-coding (in terms of isochores) sequences affect expression activity.

4. Material and methods

4.1. Data and alignment

In order to conduct our analyses, we used the publicly available RNA-seq data (Solexa/Illumina technology), produced using single-end cDNA libraries of mouse brain at stages E18 and P7 from Han et al. [25]. We aligned the reads against the mouse genome reference (UCSC release mm9; http://genome.ucsc.edu/) using REad ALigner (REAL: http://www.exelixis-lab.org/real/; [37]). REAL is based on a new, relatively simple, algorithm for the alignment of short reads onto a reference sequence. We used the appropriate arguments to allow up to two errors per read with no gaps, and to report the unique alignment (non-repeat hit) with the least number of errors. We used only the single-end data and not the paired-end produced by Han et al. [25], since REAL cannot handle paired-end reads. However, Han et al. [25] found a high consistency in their results between single-end and paired-end data (for E18 the Pearson's correlation coefficient was 0.96, and for P7 0.95), indicating that both the single-end and the paired-end sequencing yielded similar results. The results of aligning reads, for each developmental stage, to the isochores and to the coding sequences are shown in Table 1.

4.2. Expression level of isochores

To investigate the expression levels of the mouse isochores, the aligned reads were assigned to the isochores containing their mapped location. The locations and GC-levels of the isochores were extracted from [5]. To eliminate the effect of the different number of reads aligned from each developmental stage and the different length of each isochore, the aligned reads per isochore were normalized by the total count of aligned reads of the respective developmental stage and the length of the respective isochore. A scaling factor was applied at this stage, and then the $\log_2$ of each normalized read count was calculated, reflecting the expression level of each isochore. This is represented by Eq. (1), where E_L represents the expression level normalized over the length L of the isochore, R_i the number of aligned reads of the isochore, R_{ds} the total number of aligned reads for the developmental stage, and f the scaling factor.

$$E_L = \log_2\left(\frac{R_i}{R_{ds} \times L} \times f\right) \quad (1)$$

Because the normalized counts are very small, the logarithm produces negative values, however, higher expression still corresponds to peaks. Details on the isochores' coordinates, GC levels, aligned reads, and expression levels, for each developmental stage, can be found in Supplementary Table 1. As explained in the Results section, in order to account for the higher concentration of genes in isochores with higher GC level, the normalized by stage read counts

Table 1
For each developmental stage the number of aligned reads to the isochores and to the coding sequences (using the two expression thresholds).

	Development stages	
	Embryo (E18)	Postnatal (P7)
Isochores	1,643,644	1,732,507
CDSs	246,122	282,628
CDSs (≥10 aligned reads)	223,398	259,234

of each isochore were also normalized by the respective gene density, and the $\log_2$ for each isochore was calculated. If by D we denote the gene density of the isochore and by E_D the isochoric expression normalized over the gene density, Eq. (1) is modified as shown in Eq. (2).

$$E_D = \log_2\left(\frac{R_i}{R_{ds} \times D} \times f\right) \quad (2)$$

Details on the isochores' properties and their expression levels for each developmental stage can be found in Supplementary Table 2.

4.3. Expression level of genes

To investigate the expression at the gene level, the coding sequences for the mouse were retrieved from the Consensus Coding Sequence Database (CCDS; ftp://ftp.ncbi.nih.gov/pub/CCDS). From the 22,158 coding sequences (CDSs), 28 were found to lack a starting or stop codon, and were eliminated. The remaining 22,130 CDSs were assigned to isochores based on the coordinates of their exons, as given in the CCDS database. Similarly to the procedure adopted for the expression levels of the isochores, the expression level of a CDS (E_{CDS}) was produced with Eq. (3), where R_{CDS} represents the count of aligned reads in the exons of each CDS, R'_{ds} the total number of reads aligned to coding sequences for the developmental stage, and "ℓ" the length of the CDS.

$$E_{CDS} = \log_2\left(\frac{R_{CDS}}{R'_{ds} \times \ell} \times f\right) \quad (3)$$

Details on the expression levels of the CDSs, for each developmental stage, can be found in Supplementary Table 3.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2012.11.006.

Acknowledgments

We thank Prof. Giorgio Bernardi and Oliver Clay for valuable comments. SA is supported by institutional funds. KF is funded by the Greek State Scholarships Foundation. SPP is supported by the NSF-funded iPlant Collaborative (NSF grant #DBI-0735191).

References

[1] 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.
[2] G. Bernardi, Structural and Evolutionary Genomics: Natural Selection in Genome Evolution, Elsevier Science Publishers Ltd., 2005.
[3] G. Bernardi, The neoselectionist Theory of Genome Evolution, PNAS 104 (2007) 8385–8390.
[4] M. Costantini, O. Clay, F. Auletta, G. Bernardi, Isochore map of human chromosomes, Genome Res. 16 (2006) 536–541.
[5] M. Costantini, R. Cammarano, G. Bernardi, The evolution of isochore patterns in vertebrate genomes, BMC Genomics 10 (2009) 146.
[6] L. Duret, Evolution of synonymous codon usage in metazoans, Curr. Opin. Genet. Dev. 12 (2002) 640–649.
[7] O. Konu, M. Li, Correlations between mRNA expression levels and GC contents of coding and untranslated regions of genes in rodents, J. Mol. Evol. 54 (2002) 35–41.
[8] A. Vinogradov, Isochores and tissue specificity, Nucleic Acids Res. 31 (2003) 5212–5220.
[9] S. Arhondakis, F. Auletta, G. Torelli, G. D'Onofrio, Base composition and expression level of human genes, Gene 325 (2004) 165–169.
[10] J. Comeron, Selective and mutational patterns associated with gene expression in humans: influences on synonymous composition and intron presence, Genetics 167 (2004) 1293–1304.
[11] M. Semon, D. Mouchiroud, L. Duret, Relationship between gene expression and GC-content in mammals: statistical significance and biological relevance, Hum. Mol. Genet. 14 (2005) 421–427.
[12] A. Vinogradov, Dualism of gene GC content and CpG pattern in regard to expression in the human genome: magnitude versus breadth, Trends Genet. 21 (2005) 639–643.
[13] S. Arhondakis, O. Clay, G. Bernardi, Compositional properties of human cDNA libraries: practical implications, FEBS Lett. 580 (2006) 5772–5778.
[14] S. Arhondakis, O. Clay, G. Bernardi, GC level and expression of human coding sequences, Biochem. Biophys. Res. Commun. 367 (2008) 542–545.
[15] A. Mahmud, G. Amore, G. Bernardi, Compositional genome contexts affect gene expression control in sea urchin embryo, PLoS One 3 (2008) e4025.
[16] H. Caron, B. van Schaik, M. van der Mee, F. Baas, G. Riggins, P. van Sluis, M.-C. Hermus, R. van Asperen, K. Boon, P.A. Voûte, S. Heisterkamp, A. van Kampen, R. Versteeg, The human transcriptome map: clustering of highly expressed genes in chromosomal domains, Science 291 (2001) 1289–1292.
[17] R. Versteeg, B.D.C. van Schaik, M.F. van Batenburg, M. Roos, R. Monajemi, H. Caron, H.J. Bussemaker, A.H.C. van Kampen, The human transcriptome map reveals extremes in gene dentistry, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes, Genome Res. 13 (2003) 1998–2004.
[18] S. Arhondakis, K. Frousios, C.S. Iliopoulos, S.P. Pissis, G. Tischler, S. Kossida, Transcriptome map of mouse isochores, BMC Genomics 12 (2011) 511.
[19] I. Hiratani, A. Leskovar, D.M. Gilbert, Differentiation-induced replication-timing changes are restricted to AT-rich/long interspersed nuclear element (LINE)-rich isochores, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 16861–16866.
[20] L. Ren, G. Gao, D. Zhao, M. Ding, J. Luo, H. Deng, Developmental stage related patterns of codon usage and genomic GC content: searching for evolutionary fingerprints with models of stem cell differentiation, Genome Biol. 8 (2007) R35.
[21] O. Clay, G. Bernardi, GC3 of genes can be used as a proxy for isochore base composition: a reply to Elhaik et al, Mol. Biol. Evol. 28 (2011) 21–23.
[22] H. Kikuta, M. Laplante, P. Navratilova, A.Z. Komisarczuk, G.P. Engström, D. Fredman, A. Akalin, M. Caccamo, I. Sealy, K. Howe, J. Ghislain, G. Pezeron, P. Mourrain, S. Ellingsen, A.C. Oates, C. Thisse, B. Thisse, I. Foucher, B. Adolf, A. Geling, B. Lenhard, T.S. Becker, Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates, Genome Res. 17 (2007) 545–555.
[23] P. Navratilova, T. Becker, Genomic regulatory blocks in vertebrates and implications in human disease, Brief. Funct. Genomic. Proteomic. 8 (2009) 333–342.
[24] A. Sandelin, P. Bailey, S. Bruce, P.G. Engström, J.M. Klos, W.W. Wasserman, J. Ericson, B. Boris Lenhard, Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes, BMC Genomics 5 (2004) 99.
[25] X. Han, X. Wu, W.Y. Chung, T. Li, A. Nekrutenko, N.S. Altman, G. Chen, H. Ma, Transcriptome of embryonic and neonatal mouse cortex by high-throughput RNA sequencing, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 12741–12746.
[26] S. Arhondakis, S. Kossida, Compositional perspectives on human brain aging, Biosystems 104 (2011) 94–98.
[27] M. Costantini, G. Bernardi, The short-sequence designs of isochores from the human genome, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 13971–13976.
[28] A. Varriale, G. Bernardi, Distribution of DNA methylation, CpGs, and CpG islands in human isochores, Genomics 95 (2010) 25–28.
[29] S. Saccone, C. Federico, G. Bernardi, Localization of the gene-richest and the gene-poorest isochores in the interphase nuclei of mammals and birds, Gene 300 (2002) 169–178.
[30] M. Di Filippo, G. Bernardi, Mapping DNase-I hypersensitive sites on human isochores, Gene 419 (2008) 62–65.
[31] M. Di Filippo, G. Bernardi, The early apoptotic DNA fragmentation targets a small number of specific open chromatin regions, PLoS One 4 (2009) e5010.
[32] S. Arhondakis, F. Auletta, G. Bernardi, Isochores and the regulation of gene expression in the human genome, Genome Biol. Evol. 3 (2011) 1080–1089.
[33] P. Carninci, A. Sandelin, B. Lenhard, S. Katayama, K. Shimokawa, J. Ponjavic, C.A.M. Semple, M.S. Taylor, P.G. Engström, M.C. Frith, A.R.R. Forrest, W.B. Alkema, S.L. Tan, C. Plessy, R. Kodzius, T. Ravasi, T. Kasukawa, S. Fukuda, M. Kanamori-Katayama, Y. Kitazume, H. Kawaji, C. Kai, M. Nakamura, H. Konno, K. Nakano, S. Mottagui-Tabar, P. Arner, A. Chesi, S. Gustincich, F. Persichetti, H. Suzuki, S.M. Grimmond, C.A. Wells, V. Orlando, C. Wahlestedt, E.T. Liu, M. Harbers, J. Kawai, V.B. Bajic, D.A. Hume, Y. Hayashizaki, Genome-wide analysis of mammalian promoter architecture and evolution, Nat. Genet. 38 (2006) 626–635.
[34] V.B. Bajic, S.L. Tan, A. Christoffels, C. Schönbach, L. Lipovich, L. Yang, O. Hofmann, A. Kruger, W. Hide, C. Kai, J. Kawai, D.A. Hume, P. Carninci, Y. Hayashizaki, Mice and men: their promoter properties, PLoS Genet. 2 (2006) e54.
[35] A. Kel, O. Kel-Margoulis, V. Babenko, E. Wingender, Recognition of NFATp/AP-1 composite elements within genes induced upon the activation of immune cells, J. Mol. Biol. 288 (1999) 353–376.
[36] G. D'Onofrio, T.C. Ghosh, S. Saccone, Different functional classes of genes are characterized by different compositional properties, FEBS Lett. 581 (2007) 5819–5824.
[37] K. Frousios, C.S. Iliopoulos, L. Mouchard, S.P. Pissis, G. Tischler, REAL: an efficient REad ALigner for next generation sequencing reads, in: Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology, BCB'10, New York, NY, ACM, USA, 2010, pp. 154–159.