DOI:
10.1002/1873-3468.12245

Document Version
Peer reviewed version

Received Date : 29-Feb-2016

Revised Date : 27-May-2016

Accepted Date : 03-Jun-2016

Article type : Research Letter

Transcriptome activity of isochores during pre-implantation process in human and mouse.

Carl Barton[1], Costas S. Iliopoulos[2], Solon P. Pissis[2], Stilianos Arhondakis[3,4,*]

[1] *Present address:The Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK.*
[2]*Department of Informatics, King's College London, Strand, WC2R 2LS, London, UK.*
[3]*Present address:Department of Horticultural Genetics & Biotechnology, Mediterranean Agronomic Institute of Chania (MAICh), GR 73100 Chania, Crete, Greece.*
[4]*Institute of Molecular Biology and Biotechnology (IMBB), Foundation for Research and Technology – Hellas (FORTH), GR 70013 Heraklion, Crete, Greece.*

**Corresponding author*
Stilianos Arhondakis
Department of Horticultural Genetics & Biotechnology
Mediterranean Agronomic Institute of Chania (MAICh)
Alsyllio Agrokipiou, 1 Makedonias, P.O. Box 85,
Hania, 73100
Greece
Telephone: +30 2821 035000(ext. 531)
Fax: +30 2821 35001
E-mail: s.arhondakis@maich.gr

Abstract

This work investigates the role of isochores during pre-implantation process. Using RNA-seq data from human and mouse pre-implantation stages, we created the spatio-temporal transcriptional profiles of the isochores during pre-implantation. We found that from early to late stages, GC rich isochores **increase** their expression while GC-poor ones **decrease** it. **Network** analysis revealed that modules with few **co-expressed** isochores are GC-**poorer** than medium-large ones, characterized by an opposite expression as pre-implantation advances, **decreasing** and **increasing**, respectively. Our results reveal a functional contribution of the isochores, supporting the presence of structural-functional interactions during maturation and early-embryonic development.

Keywords: Base composition; GC level; expression; pre-implantation; isochores; RNA-seq; transcriptome

HIGHLIGHTS

- Isochores GC effects on their expression increases as pre-implantation advances.
- GC rich isochores increase their expression as pre-implantation advance.
- Modules with **many** co-expressed isochores **(medium/large modules)** are GC-richer and have a higher expression compared to **small** ones.
- From early to late pre-implantation stages, GC-rich modules follow an opposite expression pattern compared to the GC-poor ones, increasing and decreasing, respectively.

1. Introduction

Mammalian genomes are composed by isochores, long DNA stretches with fairly homogeneous base composition. Isochores can be further divided into a number of families characterised by different i) average GC levels, ii) gene density, iii) chromatin structure, iv) replication timing, v) gene expression, and other properties **[**reviewed in **1, 2]**. The effects of the compositional properties of coding and non-coding DNA on the expression activity and its regulatory mechanisms, is an important question in functional and evolutionary genomics.

In the last decade, several studies using genome-wide expression data corroborated a positive relationship between the GC levels of genes and their expression in samples from adult human and mouse individuals **[3-11]**. The same correlation was absent in mouse embryonic samples **[12]**, supporting the idea that low GC is fundamental during embryonic development **[13-17]**. In parallel, compositional analyses of sequences, report that GC patterns of coding sequences, promoters and/or flanking regions are major players in regulating the expression of genes and/or their functional fate in a series of pathological, e.g., cancer **[18-20]** or physiological processes, e.g., ageing **[21]**, circadian regulation **[22]**, development **[12-17]**, as also cellular functions e.g., cell-to-cell and cell-to-extracellular matrix interactions **[23-26]**. Distinct compositional properties were also found between housekeeping and tissue-specific genes **[3]**, and somatic *versus* germ-line related genes **[3, 27]**. The existence of immune cell-specific AT-rich **[28]** and CG-rich cell cycle specific promoters **[29]** has also been reported. This influence of the compositional features on the regulatory mechanisms of gene expression was recently investigated, concluding that genes embedded in distinct isochores families are subject to the control of distinct regulatory mechanisms acting at a chromatin level and of their *cis-regulatory* regions **[30].** The above could explain the detected patterns between functional classes of genes and their compositional properties, i.e., GC-rich isochores comprise more genes involved in cellular metabolism than in information storage and processing **[31].**

On the other side, embryonic development represents a highly complex process that requires the involvement of specific spatial and temporal regulations, governed by maternal transcripts and proteins. The latter, with the advance of this process begin to degrade as embryonic genome activation initiates (EGA) **[32]**. EGA reflects the period when the control of development is transferred from maternal genes to embryonic ones and it is species specific. In mouse EGA initiates at the 2-cell stage, in human and pig at the 4- to 8-cell stages, and in rabbit and bovine from 8 to 16-cell stages **[33, 34]**. EGA is characterised by structural alterations of the nuclear architecture with functional implications **[35-40]**, and it is followed by the pluripotency gene activation (PGA) **[41]**.

Here we investigate the relationship between the compositional organisation of the mammalian genomes in isochores and their implication during pre-implantation process. We aim to establish the chromosomal transcriptional landscapes of human and mouse isochores, and to assess the spatio-temporal effects of isochores' GC level on their expression activity during pre-implantation process. A network analysis is also applied in order to assess the modules of co-expressed isochores in the early development, and to investigate their structural and functional properties. Our findings promote the presence of a link between the large-scale compositional properties of the genome, the isochores, and their transcriptional activity during pre-implantation, determined by higher-order epigenetic mechanisms, like methylation, chromatin structure, known to correlate with the GC level.

2. Materials and Methods

2.1. RNA-seq data and alignment

To produce the transcriptional profiles of the mouse and human isochores we used publicly available RNA-seq data for several pre-implantation stages obtained in a recent study **[42]**. **The data consist of several replicates representing 7 and 6 pre-implantation stages of human and mouse, respectively.**The reads from each stage were aligned against the reference mouse (UCSC release mm9) and human (UCSC release hg17) genomes using REad ALigner (REAL) **[43]**. Details are given in **Supplementary Table 1.**

Supplementary Data, Tables and Figures of the manuscript can be found in the following link: http://www.inf.kcl.ac.uk/research/projects/2016_febs/

2.2. Expression level of isochores

To investigate the expression levels of the mouse and human isochores, the aligned reads were assigned to the isochores containing their mapped location (**Supplementary Table 2**). The locations and GC-spans of the isochores were extracted from **[44, 45]** for human and mouse, respectively.

To eliminate biases due to the different number of reads aligned from each sample the aligned reads per isochore (Ri) were normalised by the total count of aligned reads for each sample **(Rt). Moreover, in order to correct to the different length of each isochore the Ri was divided by** the length of each isochore. This is represented by Equation (1), where El represents the expression level normalised over the length L of the isochore, R_i the read count of the isochore, and R_t the normalised read count of the sample.

(1)
$$El_{(iso)} = \log 2(\frac{Ri}{Rt * L})$$

Summarizing, the length normalised expression metric (Equation 1) quantifies the overall transcriptional activity within each isochore, after normalization of the different length of the isochores within a same pre-implantation stage. As the normalised counts are very small the logarithm produces negative values, however, higher expression still corresponds to peaks.

In order to estimate solely the GC effects of the isochores on their expression activity, we must account for the higher concentration of genes in GC-rich isochores **[7]**. Thus, reads within each isochore were normalised by their respective gene **density (number of genes within each isochore over its length) [7]**. By D we denote the gene density of the isochore and by Ed the isochores expression normalised by the gene density, Equation (1) is modified as shown in Equation (2).

(2)
$$Ed_{(iso)} = \log 2(\frac{Ri}{Rt * D})$$

The gene density normalized expression metric (Equation 2) corrects the correlations between isochores expression and their GC by taking into account the higher amount of genes in the GC rich isochores.

The coding sequences for the mouse and human were retrieved from the Consensus Coding Sequence Database (CCDS http://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi) **[46],** and assigned to the isochores based on the coordinates of their exons as given by the CCDS database.

Details on the expression metrics for each individual sample are given in **Supplementary Tables 3** and **4**, for mouse and human respectively. Finally, in order to obtain the expression level of the isochores for each pre-implantation stage, the expression levels of the isochores for each sample (replicate) representing a same stage were averaged (human **Supplementary Table 5**, mouse, **Supplementary Table 6.**

2.3. Module networks

Module networks for the human and mouse pre-implantation stages were inferred using the LeMoNe algorithm **[47, 48]**. LeMoNe uses ensemble-based probabilistic optimisation techniques to

identify clusters of co-expressed transcripts **[49]**, in our case co-expressed isochores. The algorithm takes as input the average of the gene density normalized expression level of each isochore for each pre-implantation stage (Ed values from **Supplementary Table 5** and **6**). We identified 75 and 58 clusters of co-expressed isochores for human and mouse, respectively (**Supplementary Table 7). The network analysis aims to detected and investigate the modules of co-expressed during pre-implantation process (time-course), thus expression level of the isochores in the paternal data of human was not taken into account during modules detection.**

3. Results

3.1. The transcriptional activity of the human and mouse isochores

For each mouse and human pre-implantation stage we constructed the transcriptional landscape of the isochores across their chromosomes (**Supplementary Figure 1** and **2**). **Figure 1** shows the transcriptional landscape of the mouse and human isochores for chromosome 10 at the different stages. **Chromosome 10 was selected because it contains all isochore families and allow us to compare and contrast against our two previous studies [7, 12]. We observe that within the chromosomes the overall expression activity of the same isochore differs between different stages (see Figure 1, and Supplementary Figures 1 and 2).**

In this section, based on the herein available RNA-seq data from several pre-implantation stages we have constructed a spatio-temporal chromosomal transcriptional landscape of the mouse and human isochores during pre-implantation process (Figure 1, Supplementary Figures 1 and 2).

3.2. A time-course view of isochores' GC effects on expression activity.

The GC effects on the expression activity have been widely investigated and established in static systems, like that of single samples (tissues/cells) from adult individuals or in post-implantation developmental samples (see Introduction). In this section, we investigate the effect of isochores' GC levels upon their expression activity during the temporal and spatial process of pre-implantation. As reported elsewhere **[7, 12]**, in order to estimate solely the GC effects of the isochores' on their expression activity, we must take into account the higher gene density of the GC-rich isochores. Thus, in this section the gene density normalised expression metric was used (Equation 2 in **Material and Methods**).

For each stage we assessed the Spearman correlation between the GC level of the isochores and their expression activity (density normalized). Coefficients are given like a time-course histogram (**Figure 2;** in table format **Supplementary Table 8)**. For the human data we found positive and negative coefficients (**Figure 2A**), while for the mouse only positive ones (**Figure 2B**). In both species the coefficients were characterised by an increasing trend from early to late pre-implantation stages (Figure 2). Moreover, the weaker coefficients obtained for the early stages suggest a weaker activity of the GC rich and gene-rich isochores, can be explained by the minor gene activation and silencing of the maternal genes during early stages **[32]**, and a subsequent increased activity at the EGA and PGA stages **[41]**. **Both, EGA and PGA, can explain the differences we detected in the patterns of the compositional correlations between the two species, e.g., mouse EGA initiates at the 2-cell stage, human 4- to 8-cell stages [33, 34]. For sake of comparison, correlations are also given using the length normalized expression metric (Supplementary Table 8). As expected, correlations increase, but still an increasing trend from early to late stages can be observed.**

In order to assess those cases in which the increase of the coefficients between consecutive stages was significant, we applied a Fisher r-to-z transformation (Z scores are given in **Supplementary Table 9**). In human significance was found between the coefficients from: a) oocyte to pronuclei, b) zygote to 2-cell, and c) 4-cell to 8-cell stage (**Figure 2A**; see also **Supplementary Table 9**). In mouse significance was detected only between 2-cell and 4-cell stage (**Figure 2B; Supplementary Table 9**), which corresponds with the initiation of the EGA. Similarly, the significant increase in the coefficients between 4-cell to 8-cell stage in human coincides with the EGA (**Figure 2A**; see also **Supplementary Table 9**).

Additionally, the expression profile of each isochore family as pre-implantation advances was created (Figure 3). In human, GC-rich (H2/H3 isochores) isochores have a similar profile, while the GC-poor (L1, L2 and H1) ones have a similar pattern until the EGA initiation (4 cell stage). In mouse a similar pattern is found, although differences are less evident. Interestingly, the extreme GC rich isochore (H3) has a low expression, as that of the lowest in GC isochore family, L1. This can be explained by the near absence of H3 isochores in mouse. To this end, Ed of the oocyte was compared to that of all stages. In human, the GC-poor isochores (L1 and L2) present a statistically significant lower expression in the stages of 8 cell and morula than in the oocyte, while the GC-rich ones (H2/H3) a higher one (Supplementary Figure 3). In mouse, significance was detected only for the L2 and H1 isochores. The expression of the L2 isochore in oocyte when compared to the early stages shows a statistically significant higher expression, followed by a lower one when compared to late stages (8 cell and morula). The H1 isochore presents a statistically significant increase only

compared to the early stages, followed by an invariant expression. The above suggest that the observed shift in the correlations (Figure 2), is due to a decrease in the expression of the low GC isochores and an increase of that of the GC-rich ones during late stages (Figure 3, Supplementary table 3).

The discrepancies in coefficients, and in the number of significant differences between coefficients of adjacent stages (Figure 2, Supplementary Table 8), as also in the expression profiles between mouse and human (Figure 3, Supplementary Figure 3), can be explained, at least in part, by the less wide GC distribution of the mouse genome (lower and absent amount of L1 and H3 isochores, respectively).

In this section, we found that the solely effects of isochores' GC levels upon their expression activity is a dynamic and stage-specific phenomenon, with an increasing trend as pre-implantation advances (**Figure 2).The increasing trend of the coefficients reflects at the late stages a reduced expression activity of the low GC isochores and an increased one of the GC –rich ones (Figure 3). Interestingly, the significant differences between coefficients of adjacent stages were found to** agree with well-known and established pre-implantation events, e.g., EGA, PGA thus increasing evidence of a biological significance of the herein estimated correlations.

3.3. Isochores' co-expression network.

Genome-scale network analysis across conditions or a time course experiment is significantly important with large scale applications to biomedical research, including gene function, metabolic pathways, as well as cell function-related networks. Using the LeMoNe (Learning Module Networks) algorithm, we applied and assessed clusters of co-expressed isochores (modules) through network analysis. Very briefly, using the average density normalised expression of the human and mouse isochores, we identified clusters of co-expressed isochores (modules) across pre-implantation. As anticipated in the Material and Methods, for the human isochores we found 75 modules, while for the mouse isochores we found 58 modules (**Supplementary Table 7**).

Once the modules were established, we estimated the correlations between the number of isochores within each module and its average GC level. We found a positive correlation (human $R= 0.59$, $p< 0.0001$; mouse $R= 0.41$, $p < 0.0013$; **Figure 4**), suggesting that modules with few co-expressed isochores are GC-. Afterwards, for each pre-implantation stage we assessed the correlations between the average expression activity of each module and its average GC level. In human we found negative and statistically significant correlations, with an increasing trend as pre-

implantation advances (see **Supplementary Figure 4), indicating a decrease in the expression of the GC-poor modules. In mouse, non-significant weak positive correlations were detected (see Supplementary Figure 4). Thus, despite the increasing trend a definite conclusion could not be drawn; instead we could only suggest a possible increasing trend in the expression of the GC rich modules at late stages of mouse.**

The co-expression network analysis, **applied here for the first time upon genomic regions,** revealed existing compositional patterns between modules size (number of co-expressed isochores) and expression activity (functional property). Indeed, medium/large modules are enriched in GC-rich isochores, and have a lower expression compared to the small modules (Figure 4, Supplementary Figure 4). Moreover, we showed that from early to late pre-implantation stages, GC-poor modules decrease their expression while GC-rich ones increase it (Supplementary Figure 4). The above suggests an implication of distinct regulatory mechanisms, e.g., chromatin structure which correlate with the GC level, in the GC-rich and GC–poor modules.

4. Conclusion

This work represents an innovative exploration on the role of the isochores on the transcriptional events guiding the biological process of pre-implantation. We constructed the transcriptional landscapes of mouse and human isochores across distinct pre-implantation stages (**Figure 1**; **Supplementary Figures 1** and **2**), and showed that isochores' GC level exert a dynamic and stage-specific effect upon their transcriptional activity; favouring the expression of the GC-rich isochores in the late stages and the low GC in the early stages (**Figure 2, Figure 3, Supplementary Figure 3**). Finally, the co-expression network analysis, applied here for the first time upon the expression activity of the isochores (genomic regions), permitted to put in evidence that: i) modules with many co-expressed isochores have a rich representation of GC-richer isochores (**Figure 4**), and ii) GC-rich modules present a higher expression in late stages (Supplementary Figure 4).

Such opposite profiles between GC-rich and GC-poor modules may reflect the implication of distinct regulatory mechanisms, chromatin structure and/or DNA methylation events, modulated by their different compositional properties. Indeed, during pre-implantation a modification of the nuclear position of genes occurs, observed also in level of chromatin domains and chromosome territories (CTs) **[37-39]**. Particularly at the onset of EGA, nuclear architecture is re-organized, characterized by clustering of CTs in the nuclear periphery and the formation of a nuclear phenotype **[40]**. At the nuclear periphery, typically there is a gene-poor heterochromatin which correlates with

GC-poor isochores **[53]** and is enriched with transcriptionally silent genes, whereas the gene dense chromatin (transcriptionally active genes) resides predominantly in the nuclear interior **[54-57]**. In the context of epigenetic reprogramming during mammalian preimplantation embryos, another crucial element is DNA methylation **[58-60]**, able to reset the epigenome before pluripotency. A recent study **[61]** reported that the major event of genome-wide demethylation occurs at the 1-cell stage in mouse, and is completed at the 2-cell stage in human. In this regard, the genomic regions undergoing hypomethylation it has been shown to have to have high CpG density, while the hypermethylated ones have low CpG density the former corresponding to GC-rich regions and the latter ones to GC-poor **[62, 63**]. **Finally, despite our study did not investigated the role of the genes expression activity and their compositional properties, it worth pointing out a possible influence on the correlations of the compositionally distinct constitutive and the tissue-specific genes, GC-rich and GC-poor, respectively [3, 5]. Indeed, part of the discordant correlations and patterns between the two species may reflect the reduced GC distinction of the constitutive and tissue-specific genes in mouse, in accordance with the less pronounced isochoric structure, reflecting a less complex "epigenetic landscape" compared to human [3].**

Summarizing, epigenetic events, as those described above, modulated by the structural properties of the genome can explain the transcriptional differences we observed between GC-rich and GC-poor isochores. It is worth pointing out, that similar relationships between compositional properties of coding and non-coding DNA and the implication of distinct regulating mechanisms have been reported in post-implantation **[12-17]** and during human brain ageing **[21]**. Our results increase evidence that the regulatory mechanisms implicated during pre-implantation process are influenced by the large-scale compositional properties of the genome. This work is the first where a genome-wide compositional approach is applied to describe the transcriptional events during pre-implantation, promoting the view of the genome as a system with continuous functional and structural interactions. Future analysis using larger datasets and combining more resources may shed light on the epigenetic mechanisms that act during a biological process, influenced by the distinct compositional properties of the genome.

Acknowledgments

The authors would like to thank the anonymous reviewers for the valuable comments. We thank Giorgio Bernardi (Universita di Roma III), Oliver Clay (Universidad del Rosario, Bogota, Colombia & CIB, Medellin) and Tapash Gosh (Centre for Excellence in Bionformatics) for their valuable suggestions and constructive criticisms. Finally, we thank Fatima Vayani (Algorithms & Bioinformatics Group, King's College) for her comments and effort in editing. SA during the

preparation of this work was supported by SYNERGASIA 2009 programme from ITE/FORTH. CB is supported by an EPSRC grant (Doctoral Training Grant #EP/J500252/1). SPP is supported by a Research Grant (#RG130720) awarded by the Royal Society.

Author Contributions

SA, CSI and SPP conceived and supervised the study; SA, SPP and CB designed and performed the experiments; SA, CSI, SPP and CB wrote the manuscript. All authors approved the final version of the manuscript.

References

[1] Bernardi, G. 2005. Structural and Evolutionary Genomics: Natural Selection in Genome Evolution, Elsevier Science Publishers Ltd.
[2] Bernardi, G. 2007. The neoselectionist Theory of Genome Evolution, *PNAS*, 104, 8385-8390.
[3] Vinogradov A. 2003, Isochores and tissue specificity. *Nucleic Acids Research*. 31, 5212-5220.
[4] Comeron, J. 2004, Selective and Mutational Patterns Associated With Gene Expression in Humans: Influences on Synonymous Composition and Intron Presence. *Genetics*. 167, 1293-1304.
[5] Arhondakis, S., Auletta, F., Torelli, G., and D'Onofrio, G. 2004, Base composition and expression level of human genes. *Gene*. 325, 165-169.
[6] Arhondakis, S., Clay, O., and Bernardi, G. 2008, GC level and expression of human coding sequences. *Biochemical and Biophysical Research Communications*. 367,542-545.
[7] Arhondakis, S., Frousios, K., Iliopoulos, C.S., Pissis, S.P., Tischler, G., and Kossida, S. 2011, Transcriptome map of mouse isochores. *BMC Genomics*. 12, 511.
[8] Konu, O., and Li MD. 2002, Correlations between mRNA expression levels and GC contents of coding and untranslated regions of genes in rodents. *J Mol Evol.,* 54:35-41.
[9] Arhondakis, S., Clay, O., and Bernardi, G. 2006, Compositional properties of human cDNA libraries: Practical implications. *FEBS Letters*. 580, 5772-5778.
[10] Sabbia, V., Romero, H., Musto, H., and Naya, H. 2009, Composition profile of the human genome at the chromosome level. *J Biomol Struct Dyn*., 27, 361-370.
[11] Caron, H., van Schaik, B., van der Mee, M., et al., 2001, The Human Transcriptome Map: Clustering of Highly Expressed Genes in Chromosomal Domains. *Science*. 291, 1289-1292.
[12] Frousios, K., Iliopoulos, C.S., Tischler, G., Kossida, S., Pissis, S.P., and Arhondakis, S. 2013, Transcriptome map of mouse isochores in embryonic and neonatal cortex. *Genomics*. 101, 120-124.
[13] Hiratani, I., Leskovar, A., and Gilbert, D.M. 2004, Differentiation-induced replication-timing changes are restricted to AT-rich/long interspersed nuclear element (LINE)-rich isochores. *Proc Natl Acad Sci USA*., 101, 16861-16866.
[14] Hiratani, I., Ryba, T., Itoh, M., et al. 2008, Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol*., 6, e245.
[15] Ren., L, Gao, G., Zhao, D., Ding, M., Luo, J., and Deng, H. 2007, Developmental stage related patterns of codon usage and genomic GC content: searching for evolutionary fingerprints with models of stem cell differentiation, *Genome Biol*. 8, R35.
[16] Kikuta, H., Laplante, M., Navratilova, P., et al. 2007 Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res.,* 17, 545-555.
[17] Navratilova, P. and Becker, T.S. 2009, Genomic regulatory blocks in vertebrates and implications in human disease. *Brief Funct Genomic Proteomic*. 8, 333-342.
[18] Drucker, L., Uziel, O., Tohami, T., et al. 2003, Thalidomide down-regulates transcript levels of GC-rich promoter genes in multiple myeloma. *Mol Pharmacol*., 64, 415-420.

[19]Zhou, Y., Ma, B.G,, and Zhang, H.Y. 2007, Human oncogene tissue-specific expression level significantly correlates with sequence compositional features. *FEBS Lett.*, 581, 4361-4365.

[20]Hajjari, M., Khoshnevisan, A., and Behmanesh, M. 2014, Compositional features are potentially involved in the regulation of gene expression of tumor suppressor genes in human tissues. *Gene.* 553, 126-129.

[21]Arhondakis, S., and Kossida, S. 2011, Compositional perspectives on human brain aging. *Biosystems*. 104, 94-98.

[22]Bozek, K., Relógio, A., Kielbasa, S.M., et al. 2009, Regulation of clock-controlled genes in mammals. *PLoS One.* 4, e4882

[23]Cooke, D.W., Bankert, L.A., Roberts, C.T., Jr, LeRoith, D., and Casella, S.J. 1991, Analysis of the human type I insulin-like growth factor receptor promoter region. *Biochem Biophys Res Commun*., 177, 1113-1120.

[24]Hotta, H., Miyamoto, H., Hara, I., Takahashi, N., and Homma, M. 1992, Genomic structure of the ME491/CD63 antigen gene and functional analysis of the 5'-flanking regulatory sequences. *Biochem Biophys Res Commun.,* 185, 436-442.

[25]Villa-Garcia, M., Li, L., Riely, G., and Bray, P.F. 1994, Isolation and characterization of a TATA-less promoter for the human beta 3 integrin gene. Blood. 83, 668-676.

[26]Bajic, V.B., Tan, S.L., Christoffels, A., et al. 2006 Mice and men: their promoter properties, *PLoS Genet.,* 2, e54.

[27]Vinogradov, A. 2005, Dualism of gene GC content and CpG pattern in regard to expression in the human genome: Magnitude versus breadth. *Trends in Genetics*. 21, 639-643.

[28]Kel, A., Kel-Margoulis, O., Babenko, V., and Wingender, E. 1999, Recognition of NFATp/AP-1 composite elements within genes induced upon the activation of immune cells. *J. Mol. Biol.,* 288, 353–376.

[29]Kel, A.E., Kel-Margoulis, O.V., Farnham, P.J., Bartley, S.M., Wingender, E., and Zhang, M.Q. 2001, Computer-assisted identification of cell cycle-related genes: new targets for E2F transcription factors. *J Mol Biol.*, 309, 99-120.

[30]Arhondakis, S., Auletta, F., and Bernardi, G. 2011, Isochores and the regulation of gene expression in the human genome. *Genome Biol Evol.,* 3, 1080-1089.

[31]D'Onofrio, G., Ghosh, T.C., and Saccone, S. 2007, Different functional classes of genes are characterized by different compositional properties. *FEBS Lett.,* 581, 5819–5824.

[32]Tadros, W., and Lipshitz, H.D. 2009, The maternal-to-zygotic transition: a play in two acts. *Development.* 136, 3033–3042.

[33]Sirard, M.A. 2012, Factors affecting oocyte and embryo transcriptomes. *Reprod Domest Anim.,* 47, 148–155.

[34]Braude, P., Bolton, V., and Moore, S. 1988, Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. *Nature* 332, 459–461.

[35]Cremer, T., and Zakhartchenko, V. 2011, Nuclear architecture in developmental biology and cell specialisation. *Reprod Fertil Dev.,* 23, 94-106.

[36]Li, L., Lu, X., and Dean, J. 2013, The maternal to zygotic transition in mammals. *Mol Aspects Med.,* 34, 919-938;

[37]Shi, W., Zakhartchenko, V., and Wolf, E. 2003, Epigenetic reprogramming in mammalian nuclear transfer. *Differentiation* 71, 91-113.

[38]Koehler, D., Zakhartchenko, V., Froenicke, L., et al. 2009, Changes of higher order chromatin arrangements during major genome activation in bovine preimplantation embryos. *Exp Cell Res.,* 315, 2053-2063.

[39]Mehta, I.S., Eskiw, C.H., Arican, H.D., Kill, I.R., and Bridger, J.M. 2011, Farnesyltransferase inhibitor treatment restores chromosome territory positions and active chromosome dynamics in Hutchinson-Gilford progeria syndrome cells. *Genome biology*. 12, R74

[40]Solovei, I., Kreysing, M., Lanctot, C., et al. 2009, Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell* 137, 356-368.

[41]Popken, J., Koehler, D., Brero, A., et al. 2014, Positional changes of a pluripotency marker gene during structural reorganization of fibroblast nuclei in cloned early bovine embryos. *Nucleus*. 5, 542-554.

[42]Xue, Z., Huang, K., Cai, C., et al. 2013, Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* 500, 593-597.

[43]Frousios, K., Iliopoulos, C.S., Mouchard. L, et al: 2010, REAL: an efficient Read ALigner for next generation sequencing reads. Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology New York, NY, USA: ACM, 154-159.

[44]Costantini, M., Clay, O., Auletta, F., and Bernardi, G. 2006, Isochore Map of Human Chromosomes. *Genome Research* 16, 536-541.

[45]Costantini, M., Cammarano, R., and Bernardi, G. 2008, The evolution of isochore patterns in vertebrate genomes. *BMC Genomics* 10, 146

[46]Pruitt, K.D., Harrow, J., Harte, R.A., et al. 2009 The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res.,* 19, 1316-1323.

[47]Bonnet, E, Tatari, M., Joshi, A., et al 2010, Module network inference from a cancer gene expression data set identifies microRNA regulated modules. *PloS one*, 5: e10162.

[48]Cramer, G.R., Urano, K., Delrot, S., Pezzotti, M. and Shinozaki, K. 2011, Effects of abiotic stress on plants: a systems biology perspective. *BMC plant biology* 11, 163.

[49]Bonnet, E., Michoel, T., and Van de Peer, Y. 2010, Prediction of a gene regulatory network linked to prostate cancer from gene expression, microRNA and clinical data. Bioinformatics (Oxford, England) 26: i638–44.

[50]Zhang, L., and Li, W.H. 2004, Mammalian housekeeping genes evolve more slowly than tissue-specific genes. *Mol Biol Evol.* 21, 236-239.

[51]Kommadath, A., Nie, H., Groenen, M.A., Pas, M.F., Veerkamp, R.F., and Smits, M.A. 2011, Regional regulation of transcription in the bovine genome. *PLoS One.* 6, e20413.

[52]Nie, H., Crooijmans, R.P., Lammers, A., et al. 2010 Gene expression in chicken reveals correlation with structural genomic features and conserved patterns of transcription in the terrestrial vertebrates. *PLoS One*. 5, e11990.

[53]Saccone, S., Federico, C. and Bernardi, G. 2002, Localization of the gene-richest and the gene-poorest isochores in the interphase nuclei of mammals and birds. *Gene* 300, 169-178.

[54]Bolzer, A., Kreth, G., Solovei, I., et al. 2005, Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. PLoS Biol 3, e157.

[55]Fedorova, E., and Zink, D. 2009, Nuclear genome organization: common themes and individual patterns. *Curr Opin Genet Dev.,* 19, 166-171.

[56]Küpper, K., Kolbl, A., Biener, D., 2007, Radial chromatin positioning is shaped by local gene density, not by gene expression. Chromosoma 116, 285-306.

[57]Schneider, R., and Grosschedl, R. 2007, Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *Genes Dev.,* 21, 3027-3043.

[58] Bird, A. (2002) DNA methylation patterns and epigenetic memory. Genes Dev. 16, 6–21

[59] Feng, S., Jacobsen, SE., Reik, W. (2010) Epigenetic reprogramming in plant and animal development. Science 330, 622–627.

[60] Fulka, H., Mrazek, M., Tepla, O. & Fulka, J. Jr. DNA methylation pattern in human zygotes and developing embryos. Reproduction 128, 703–708 (2004).

[61] Guo H, Zhu P, Yan L, et al., (2014) The DNA methylation landscape of human early embryos. Nature. 511:606-610.

[62] Fenouil R, Cauchy P, Koch F, et al. (2012) CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. Genome Research. 22:2399-2408.

[63] Varriale A, Bernardi G. (2010) Distribution of DNA methylation, CpGs, and CpG islands in human isochores. Genomics. 95:25-28.

Figure Legends

Figure 1: Compositional overview of the human (upper panel) and mouse (lower panel) chromosome 10 (Mb) and the corresponding expression profile of the available samples. The color code spans the spectrum of GC levels in the isochore families, indicated from ultramarine blue (GC-poorest L1 isochore family) to scarlet red (GC-richest H3 isochore family). *Y-axis* denotes the expression levels (EL - Equation (1)) transformed in positive values. High expression corresponds to peaks.

Figure 2: The profile of the spearman correlation coefficients between expression of the isochores (gene density normalised- Equation (2)) and their GC level as pre-implantation advances for human (Panel A) and mouse (Panel B). Arrows indicate the direction (increase or decrease) of those cases where the difference in coefficients between adjacent stages was significant (*p-values* are given in Supplementary Table 8, panel A and C). Each colour indicates a specific stage.

Figure 3: The profile of the average gene density normalized expression activity (Ed) of each isochore family during pre-implantation process in human (Upper panel) and mouse (lower panel).

Figure 4: The correlations between the number of the isochores within each module and their corresponding average GC level. The upper panel reports the correlations for the human modules, and the lower panel for the mouse modules.

Supplementary Figures

Supplementary Figure 1: Compositional overview of the human chromosomes (Mb) and their corresponding expression profiles for each available sample. The color code spans the spectrum of GC levels in the isochore families, indicated from ultramarine blue (GC-poorest L1 isochore family) to scarlet red (GC-richest H3 isochore family). *Y-axis* denotes the expression levels (EL - Equation (1)) transformed in positive values. High expression corresponds to peaks.

Supplementary Figure 2: Compositional overview of the mouse chromosomes (Mb) and their corresponding expression profiles for each available sample. The color code spans the spectrum of GC levels in the isochore families, indicated from ultramarine blue (GC-poorest L1 isochore family) to scarlet red (GC-richest H3 isochore family). *Y-axis* denotes the expression levels (EL - Equation (1)) transformed in positive values. High expression corresponds to peaks.

Supplementary Figure 3: **The p-values (Cochran test, non-parametric t-test) between the average expression level of each isochore in the oocyte stage compared to that of each stage. Upper panel for human, and lower panel for mouse. Green cells indicate a higher expression in that specific stage compared to oocyte (increase) and the red ones a lower one (decrease).**

Supplementary Figure 4: The spearman correlations between the average GC level (%) of each module and its corresponding average expression level, for each stage. On the left in table format and on the right in histogram. Upper panel for human and lower for mouse.

Supplementary Tables

Supplementary Table 1: This table reports for each human and mouse sample, the GSM accession (Sample record), the description, and the SRA (Sequence Read Archive) accession, and the number of raw reads. Moreover, report the aligned reads per sample in number and in percentage (%).

Supplementary Table 2: This table reports for each isochore the number of aligned reads per isochores. The worksheet "Aligned Reads Human" contains the aligned reads for the human isochores, and the worksheet "Aligned Reads Mouse" for the mouse isochores.

Supplementary Table 3: This table reports the expression level of each human isochore for all available samples. The worksheet (El_Hs) reports the length normalised expression (El) of the human isochores, and the worksheet (Ed_Hs) the gene density normalised expression (Ed) of the human isochores.

Supplementary Table 4: This table reports the expression level of each mouse isochore for all available samples. The worksheet (El_Mm) reports the length normalised expression (El) of the mouse isochores, and the worksheet (Ed_Mm) the gene density normalised expression (Ed) of the mouse isochores.

Supplementary Table 5: This table reports the average expression level of each human isochore for the samples representing a same stage. The worksheet (El_Hs) reports the average of the length normalised expression metric (El) of the human isochores, and the worksheet (Ed_Hs) the average of the gene density normalised (Ed).

Supplementary Table 6: This table reports the average expression level of each mouse isochore for the samples representing a same stage. The worksheet (El_Mm) reports the average of the length normalised expression metric (El) of the mouse isochores, and the worksheet (Ed_Mm) the average of the gene density normalised (Ed).

Supplementary Table 7: This table reports information for the human and mouse modules. The first column reports the isochore, the 2nd the module in which the isochore belongs, the GC % level of the isochore, and the corresponding expression in the different stages (gene density normalised expression metric). The last columns indicate the isochore family. The worksheet "Modules Human" reports human modules, and the worksheet "Modules mouse" for mouse.

Supplementary Table 8: This table reports for each sample the coefficients and the corresponding p-values, as estimated between isochores' GC and their corresponding expression activity. Panels A and C, report the spearman correlations when the gene density normalised expression metric was used (Ed), while B and D when the length normalised (El) was used instead. Panel A and B are for human, and C and D for mouse.

Supplementary Table 9: This table reports the Fisher r-to-z transformation results applied upon coefficients of consecutive stages. Z scores and p-values are given. Green cells denote those significant cases.

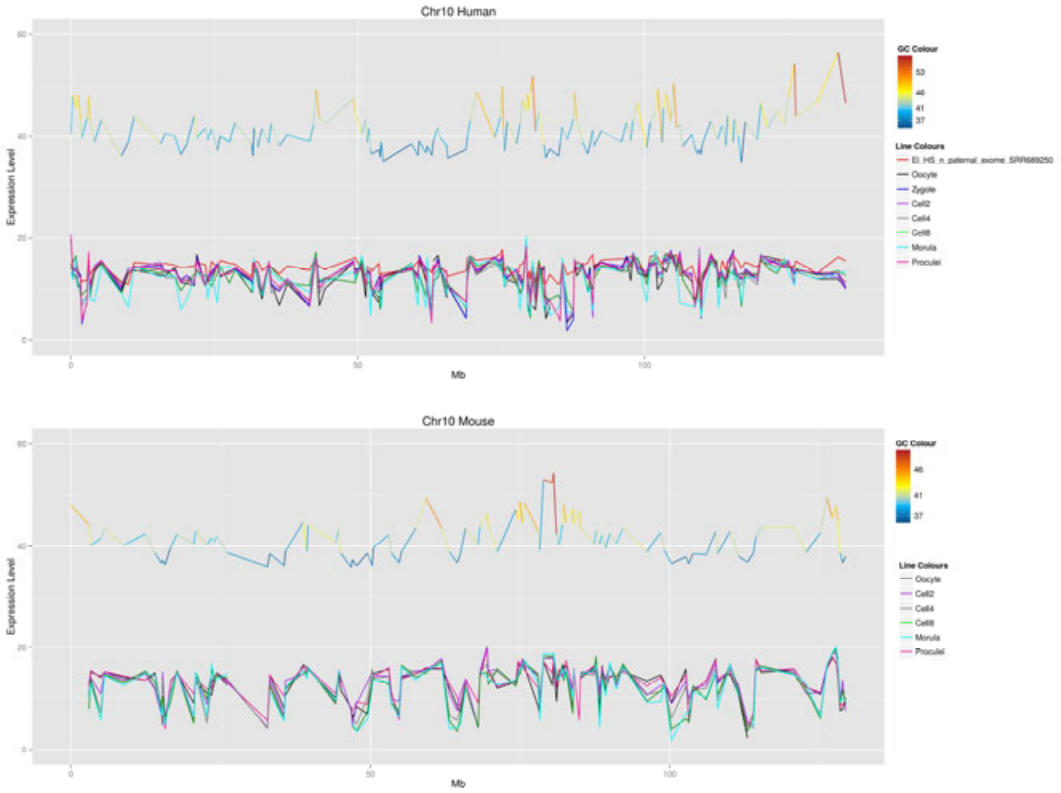

Fig. 2