Update

Genome Analysis

Monoallelic expression and tissue specificity are associated with high crossover rates

Anamaria Necsulea[1], Marie Sémon[2], Laurent Duret[1] and Laurence D. Hurst[3]

[1] Université de Lyon, F-69000, Lyon; Université Lyon 1; CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France; HELIX, Unité de recherche INRIA
[2] Institut de Génomique Fonctionnelle de Lyon, CNRS UMR5242 - INRA 1288-ENS-UCBL, IFR128 Biosciences Lyon-Gerland, Ecole Normale Supérieure de Lyon
[3] Department of Biology and Biochemistry, University of Bath, Somerset BA2 7AY, UK

What determines the recombination rate of a gene? Following the observation that, in humans, imprinted genes have unusually high recombination levels, we ask whether increased recombination is seen for other monoallelically expressed genes and, more generally, how transcriptional properties relate to recombination. We find that monoallelically expressed genes do have high crossover rates and discover a striking negative correlation between within-gene crossover rate and expression breadth. We hypothesise that these findings are possibly symptomatic of a more general, adverse relationship between recombination and transcription in the human genome.

Introduction

Recombination rates are highly heterogeneous in the human genome, but the underlying reasons are not transparent. The finding that imprinted domains have unusually high crossover rates [1,2] suggests a possible coupling between recombination and transcriptional properties of a gene. Here, we follow up this initial observation to ask whether high crossover rates are found for monoallelically expressed genes more generally and, in addition, whether the breadth of expression of a gene might predict its crossover rate.

Parental imprinting and meiotic recombination

Parental imprinting is generally understood as a form of epigenetic regulation of gene expression: for imprinted genes, only one allele is transcribed, and the choice of the transcribed allele is dependent on its parent of origin. While this definition reduces the notion of imprint to its impact on transcription, other fundamental cellular mechanisms are affected by parental imprinting, and the evolutionary persistence of this phenomenon might be explained, in part, by these often ignored factors [3]. In this respect, meiotic recombination is perhaps the most relevant process, as there is evidence for significant parent-of-origin effects on crossover rates [4]. Strikingly, the only known occurrence of an imprinted recombination hotspot coincides with a transcriptionally imprinted domain [5].

These recent findings strengthen the existing evidence in favour of an association between imprinting and meiotic recombination. Transcriptionally imprinted regions have two other particular characteristics with respect to recombination: they display strong differences between male and female crossover rates (heterochiasmy) [2,6,7] and their sex-averaged crossover rates are unusually high [1,2].

Before attempting to propose a biological interpretation for the association between transcriptional imprinting and meiotic recombination, the strength of this association must be rigorously assessed. In this study, we focus on the unusually high sex-averaged crossover rates observed for imprinted genes. To test whether this property is uniquely associated with parental imprinting, we analyse an alternative hypothesis: that high crossover rates might be a general feature of genes with monoallelic expression. This hypothesis is plausible, because some features of imprinted loci, such as asynchronous DNA replication [8], are encountered also for the randomly inactivated X chromosome in females [9] and for genes undergoing random allelic exclusion [10].

Here, we analyse human autosomal genes with haploid expression, taking advantage of the recent identification of the first large-scale dataset of genes subject to random monoallelic expression [11]. We show that high levels of crossover are a general property of autosomal genes with monoallelic expression and, unexpectedly, we find that the within-gene crossover rate is negatively correlated with the gene expression breadth. We suggest that a common process that associates increased recombination with decreased transcription levels might explain these observations.

High crossover rate for autosomal genes with monoallelic expression

We identified 51 human genes with strong evidence for parental imprinting, using two online resources: http://igc.otago.ac.nz [12] and http://geneimprint.org. In addition, we analysed 3423 genes with biallelic expression (BE genes) and 357 genes with random monoallelic expression (RME genes), identified in human B-lymphoblastoid cells [11]. For the latter class of genes, the choice of the expressed allele is independent of the parent of origin, varies between cell lines, and some cells can express both alleles [11].

We calculated the within-gene crossover rate using fine-scale genetic maps built from single nucleotide polymorphism data [13]. As expected, imprinted genes have very high

Corresponding author: Necsulea, A. (necsulea@biomserv.univ-lyon1.fr); Hurst, L.D. (l.d.hurst@bath.ac.uk).

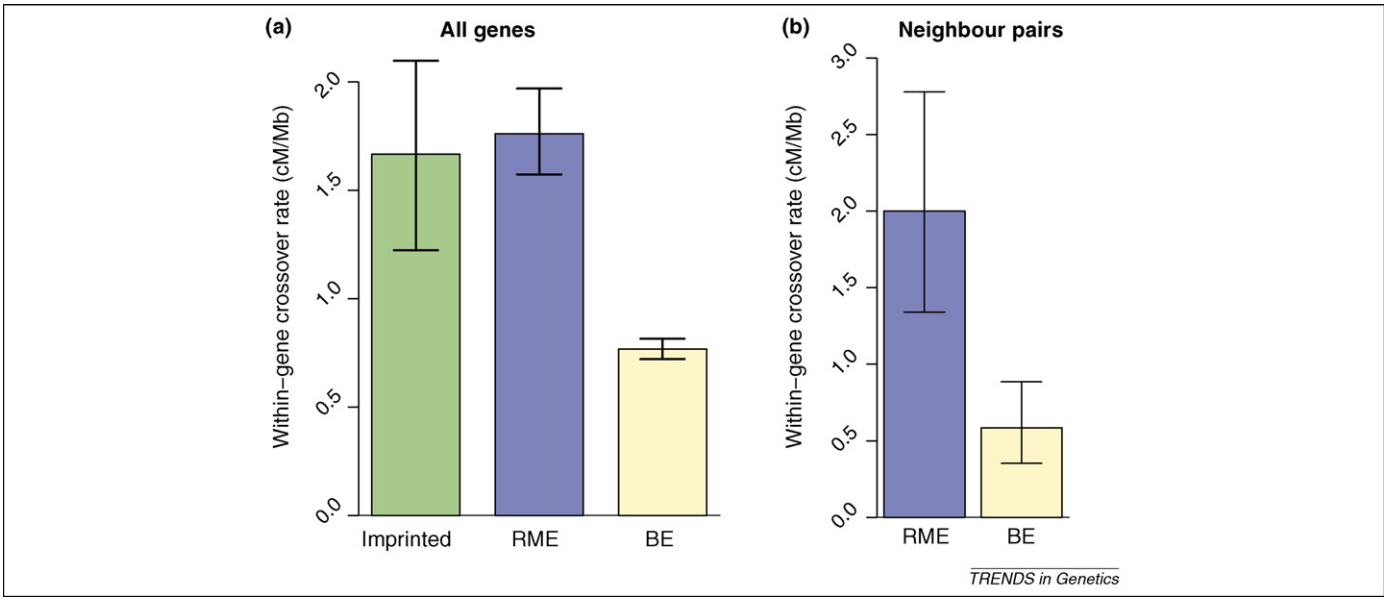

Figure 1. (**a**) Mean and confidence intervals (computed with randomizations) of the within-gene crossover rates observed for imprinted, RME and BE genes. (**b**) Mean and confidence intervals of the within-gene crossover rates for 65 pairs of BE, RME neighbour genes, separated by less than 25 kb.

crossover rates, with a mean at 1.67 cM/Mb, higher than that of non-imprinted genes (0.86 cM/Mb). We tested the difference in mean crossover rates using a randomization procedure (see methods in supplementary material online) and found that it is highly significant ($P < 10^{-3}$); a comparison of the median rates with the Wilcoxon test yielded the same conclusion (medians 1.09 and 0.25 cM/Mb, respectively, $P = 1.97 \times 10^{-5}$).

Strikingly, RME genes also have significantly higher crossover rates (mean 1.76 cM/Mb) than BE genes (mean 0.77 cM/Mb, $P < 10^{-3}$, Figure 1a). RME and imprinted genes are not significantly different ($P = 0.62$). The same conclusions were reached when measuring crossover rates in fixed-size windows, regardless of their position within the gene (Table S1 and Figure S1 in the supplementary material online). The difference between RME and BE genes remains striking even when comparing pairs of (RME, BE) neighbour genes separated by less than 25 kb (Figure 1b). This suggests considerable specificity of the effect and indicates that genomic location cannot explain the discrepancy in crossover rates. Furthermore, the distance between crossover hotspots and transcription start sites is significantly smaller for imprinted and RME genes (medians 13.2 kb and 12.6 kb) than for BE genes (median 29.0 kb, Wilcoxon test, $P < 10^{-5}$).

Crossover rates are negatively correlated with gene expression breadth

We next wanted to test whether other aspects of gene expression correlate with crossover rates. Prompted by the report that genes with random monoallelic expression are often tissue-specific [11], we analysed the relationship between crossover rates and gene expression breadth, estimated with EST, SAGE and microarray data (see methods in supplementary material online). We find that the within-gene crossover rate is significantly and negatively correlated with the gene expression breadth (Figure 2a, Spearman's ρ −0.23, −0.15 and −0.12 with EST, SAGE and microarray data, respectively, $P < 10^{-10}$).

Could the correlation between expression breadth and crossover rates explain our observations for monoallelically transcribed genes? This possibility cannot be *a priori* excluded, because imprinted and RME genes are expressed in a narrower range of tissues (medians 14 and 17 tissues with EST data, respectively) than BE genes (median 27 tissues, Wilcoxon test, $P < 10^{-3}$). We developed a randomization procedure to test whether monoallelic transcription remains significantly associated with high crossover rates when controlling for expression breadth: we draw 1000 subsets of BE genes that have the same expression breadth distribution as imprinted or RME genes, and we compare the mean crossover rate of each subset with that observed for imprinted or RME genes (methods). In most cases, these subsets of BE genes have higher mean crossover rates than the whole BE dataset, as expected given the correlation between expression breadth and crossover rates (Figure 2b). However, in more than 95% of the cases, the BE subsets have lower rates than imprinted and RME genes (Figure 2b; Table S2 in the supplementary material online), indicating that expression breadth is not sufficient to explain the difference in crossover rates between genes with monoallelic and biallelic expression.

Robustness of the association between transcription patterns and crossover rates

Genes with monoallelic and biallelic expression can differ in aspects other than their transcription pattern, and so can broadly expressed and tissue-specific genes. Here, we ask if the association between transcription patterns and crossover rates can be explained by other characteristics of the genes involved. We have analysed a wide range of genomic features, such as the presence of CpG island promoters, the frequency of repeated elements in intronic regions, the gene length, the exonic fraction, the GC-content (Table S1 and the supplementary material online). In

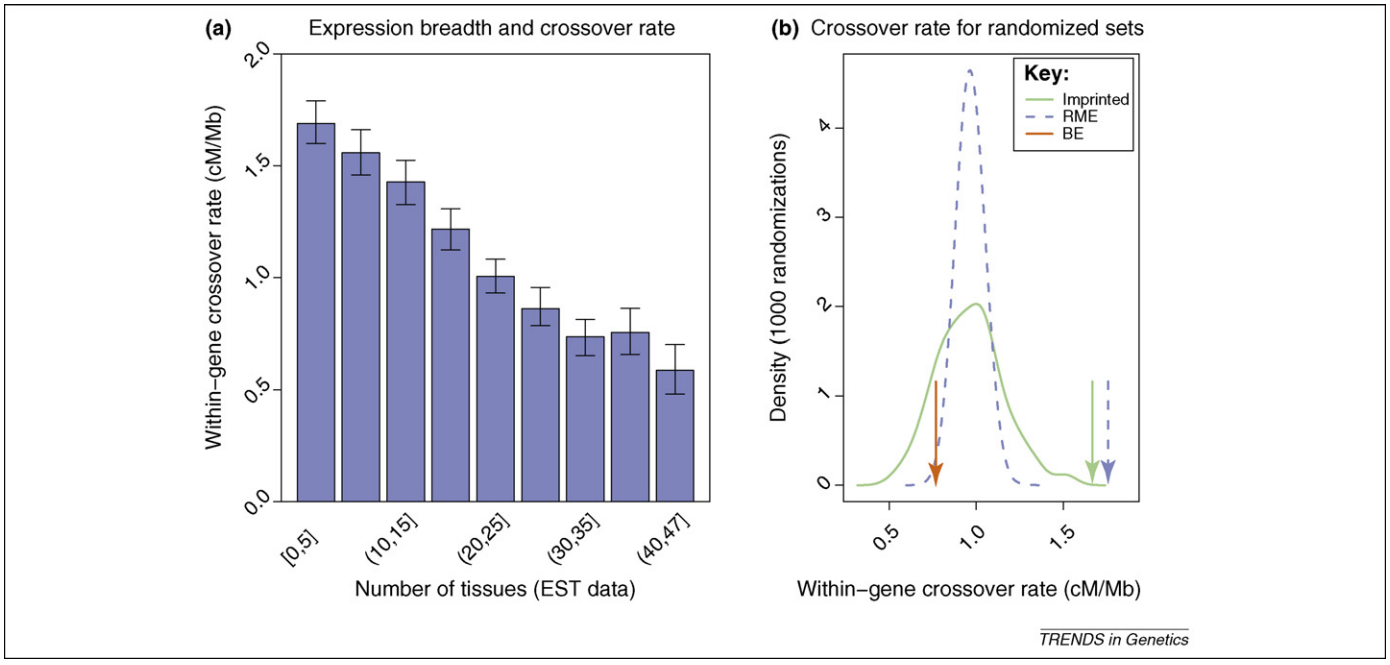

Figure 2. (**a**) Variation of within-gene crossover rate as a function of expression breadth, measured with EST data. (**b**) Distribution of mean within-gene crossover rates for randomized datasets of genes with biallelic expression (density curves). The arrows represent the observed means for imprinted genes (unbroken green line), RME genes (broken blue line) and BE genes (unbroken orange line). The two density curves represent two distinct randomizations: in the first, the expression breadth distribution of the randomized subsets is identical with that observed for imprinted genes (unbroken green line) and in the second, to that observed for RME genes (broken blue line).

particular, we controlled for the presence of sequence motifs associated with recombination hotspots [14,15] and for the distance to telomeres, known to influence the crossover rate [16]. The difference in crossover rates between RME and BE genes remains statistically significant when accounting for these factors, and so does the correlation between expression breadth and crossover rates (Figures S12–S24, Tables S4–S9 and the supplementary material online). However, for the comparison between imprinted and BE genes, the difference is not always statistically significant, probably because the dataset available for imprinted genes is small.

Crossover rates can be influenced by epigenetic factors, such as DNA methylation, in addition to sequence-encoded characteristics. It was reported that the level of polymorphism on CpG dinucleotides, which can be used as an indirect measure of germline DNA methylation, is positively correlated with the crossover rate [17]. Here, we test whether this factor can explain the relationships that we observed between transcription patterns and crossover rates. To control for variations in polymorphism due to other factors (such as hitch-hiking or background selection), we analyse the ratio r of the level of CpG polymorphism, normalized by the level of non-CpG polymorphism, computed in intronic regions (methods in supplementary material online).

RME and tissue-specific genes generally have significantly higher values of r than BE and broadly expressed genes, respectively (Wilcoxon test, $P < 10^{-10}$; Table S1 and Figures S23 and S24 in the supplementary material online). However, we find that the difference between RME and BE genes and the correlation between the expression breadth and crossover rates remain significant when controlling for this factor (Figures S25, S26 and the supplementary material online). We can thus conclude that the level of germline DNA methylation is not sufficient to explain the relationships that we observed between transcription patterns and crossover rates.

A previously unknown link between recombination and transcription in the human genome

We present two intriguing observations that suggest a connection between transcription and recombination in the human genome: first, that genes with monoallelic expression (and not just imprinted genes) have significantly higher within-gene crossover rates than biallelically transcribed genes, and second, that the tissue expression breadth is negatively correlated with the within-gene crossover rate. Importantly, both observations remain valid when controlling for numerous potential confounding factors, suggesting that the transcription pattern is genuinely associated with the rate of crossover within genes.

We show that these two properties are independent: when controlling for expression breadth variation, the discrepancy between monoallelic and biallelic genes remains significant. It does not necessarily follow that the underlying biological mechanisms are also independent: these observations might be two facets of the same process that (directly or indirectly) ties together increased recombination and reduced transcription. If there is indeed a unique mechanistical cause for this association between transcription and recombination, a simple prediction can be made: crossover rates should also be negatively correlated with aspects of gene transcription other than tissue expression breadth and monoallelic/biallelic expression. An appealing perspective that ensues from this work is to study the relationship between the crossover rate and the expression level in meiotic cells, where recombination occurs. In somatic tissues, monoallelic and tissue-specific

genes have lower transcription levels than biallelic and broadly expressed genes (see the online supplementary material). If this is also the case for the germline, and especially for meiotic cells, then the expression level in the latter might provide the key towards understanding the underlying biological mechanisms.

Discovering the existence of a correlation between meiotic crossover and transcription patterns is in itself striking, but perhaps the most surprising element here is the direction of the correlation. Although no direct relationship between transcriptional activity and meiotic recombination has been reported, the occurrence of crossovers is known to be favoured by the open chromatin at promoters [18], and by the binding of certain transcription factors [19]. From these observations, one might intuitively predict a positive correlation between crossover rates and transcription, and this assumption lingers in the literature [20,21], despite the paucity of supporting evidence. Our findings contradict this common assumption, and provide support for a negative association between recombination and transcription in humans. Whether the same is true for other eukaryotic species, and what might be the underlying cause, remain for now open questions.

Acknowledgements

This work was funded by the Centre National de la Recherche Scientifique, the Agence Nationale de la Recherche (ANR-08-GENO-003-01) and by the European Community Framework 6 Programme, via the coordination action Integrating Cooperation Research Across Europe (INCORE); a visiting fellowship awarded to A.N. L.D.H. is a Royal Society Wolfson Research Merit Award Holder. We thank Henrik Kaessmann and Thomas Lenormand for helpful suggestions.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tig.2009.10.001.

References

1 Lercher, M.J. and Hurst, L.D. (2003) Imprinted chromosomal regions of the human genome have unusually high recombination rates. *Genetics* 165, 1629–1632

2 Sandovici, I. *et al.* (2006) Human imprinted chromosomal regions are historical hot-spots of recombination. *PLoS Genet.* 2, e101

3 Pardo-Manuel de Villena, F. *et al.* (2000) Natural selection and the function of genome imprinting: beyond the silenced minority. *Trends Genet.* 16, 573–579

4 Paigen, K. *et al.* (2008) The recombinational anatomy of a mouse chromosome. *PLoS Genet* 4, e1000119

5 Ng, S. *et al.* (2009) Parental origin of chromosomes influences crossover activity within the *Kcnq1* transcriptionally imprinted domain of *Mus musculus*. *BMC Mol. Biol.* 10, 43

6 Paldi, A. *et al.* (1995) Imprinted chromosomal regions of the human genome display sex-specific meiotic recombination frequencies. *Curr. Biol.* 5, 1030–1035

7 Robinson, W.P. and Lalande, M. (1995) Sex-specific meiotic recombination in the Prader–Willi/*Angelman syndrome* imprinted region. *Hum. Mol. Genet.* 4, 801–806

8 Kitsberg, D. *et al.* (1993) Allele-specific replication timing of imprinted gene regions. *Nature* 364, 459–463

9 Schmidt, M. and Migeon, B.R. (1990) Asynchronous replication of homologous loci on human active and inactive X chromosomes. *Proc. Natl Acad. Sci. USA* 87, 3685–3689

10 Mostoslavsky, R. *et al.* (2001) Asynchronous replication and allelic exclusion in the immune system. *Nature* 414, 221–225

11 Gimelbrant, A. *et al.* (2007) Widespread monoallelic expression on human autosomes. *Science* 318, 1136–1140

12 Morison, I.M. and Reeve, A.E. (1998) A catalogue of imprinted genes and parent-of-origin effects in humans and animals. *Hum. Mol. Genet.* 7, 1599–1609

13 The International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, 851–861

14 Myers, S. *et al.* (2005) A fine-scale map of recombination rates and hotspots across the human genome. *Science* 310, 321–325

15 Myers, S. *et al.* (2008) A common sequence motif associated with recombination hot spots and genome instability in humans. *Nat. Genet.* 40, 1124–1129

16 Kong, A. *et al.* (2002) A high-resolution recombination map of the human genome. *Nat. Genet.* 31, 241–247

17 Sigurdsson, M.I. *et al.* (2009) HapMap methylation-associated SNPs, markers of germline DNA methylation, positively correlate with regional levels of human meiotic recombination. *Genome Res.* 19, 581–589

18 Wu, T.C. and Lichten, M. (1994) Meiosis-induced double-strand break sites determined by yeast chromatin structure. *Science* 263, 515–518

19 White, M.A. *et al.* (1993) Transcription factors are required for the meiotic recombination hotspot at the *HIS4* locus in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* 90, 6621–6625

20 Vinogradov, A.E. (2003) Isochores and tissue-specificity. *Nucleic Acids Res.* 31, 5212–5220

21 Dreszer, T.R. *et al.* (2007) Biased clustered substitutions in the human genome: the footprints of male-driven biased gene conversion. *Genome Res.* 17, 1420–1430

0168-9525/$ – see front matter.

doi:10.1016/j.tig.2009.10.001 Available online 21 October 2009