An Integrative Breakage Model of genome architecture, reshuffling and evolution

The *Integrative Breakage Model* of genome evolution, a novel multidisciplinary hypothesis for the study of genome plasticity

Marta Farré[1)†], *Terence J. Robinson*[2)] *and Aurora Ruiz-Herrera*[1)3)*]

Our understanding of genomic reorganization, the mechanics of genomic transmission to offspring during germ line formation, and how these structural changes contribute to the speciation process, and genetic disease is far from complete. Earlier attempts to understand the mechanism(s) and constraints that govern genome remodeling suffered from being too narrowly focused, and failed to provide a unified and encompassing view of how genomes are organized and regulated inside cells. Here, we propose a new multidisciplinary *Integrative Breakage Model* for the study of genome evolution. The analysis of the high-level structural organization of genomes (nucleome), together with the functional constrains that accompany genome reshuffling, provide insights into the origin and plasticity of genome organization that may assist with the detection and isolation of therapeutic targets for the treatment of complex human disorders.

Keywords:
epigenome; evolution; genome reshuffling; nucleome; recombination; selection

DOI 10.1002/bies.201400174

[1)] Departament de Biologia Cel·lular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona, Campus UAB, Barcelona, Spain
[2)] Evolutionary Genomics Group, Department of Botany and Zoology, University of Stellenbosch, Stellenbosch, South Africa
[3)] Genome Integrity and Instability Group, Institut de Biotecnologia i Biomedicina (IBB), Universitat Autònoma de Barcelona, Campus UAB, Barcelona, Spain

[†]Present address: Department of Comparative Biomedical Sciences, The Royal Veterinary College, London, United Kingdom.

***Corresponding author:**
Aurora Ruiz-Herrera
E-mail: aurora.ruizherrera@uab.cat

Abbreviations:
CTCF, CCCTC-binding factor; **CNV,** copy number variant; **DHSs,** DNase I hypersensitive sites; **DSB,** double strand break; **EBR,** evolutionary breakpoint region; **GR,** genomic rearrangement; **SD,** segmental duplication; **TADs,** topological associated domains; **TE,** transposable element; **TR,** tandem repeat.

Introduction

How genomes are organized and restructured through genomic rearrangements (GRs) are questions of fundamental importance for understanding the dynamics of chromosomal evolution, the evolutionary relationships among species, and in the longer term, speciation. A useful analogy for the striking effect of structural modification is provided by Peng et al. [1]: ''...if a genome is compared to a continental landform, then one type of change—point mutations—is analogous to gradual changes in the landscape due to erosion by wind and water. A second type of change—genome rearrangements—comprises evolutionary ''earthquakes'' that dramatically change the landscape.'' GRs, initially caused by double strand breaks (DSBs), can cause lethal genetic alterations resulting in cell death or new variants that enhance genome instability. However, these new chromosomal forms can have important heritable implications if they arise in the early stages of embryonic development, or in the germ line. They can result in structural rearrangements that are transmitted either as potentially deleterious genetic anomalies, or as new chromosomal variants that are associated with some selective advantage that possibly facilitates speciation.

Despite an extensive literature, the contrasting behaviors of GRs (on the one

hand an evolutionary marker and on the other a potentially life-threatening neoplasm) are far from being fully understood. However, the increasing availability of high-throughput techniques offer unprecedented ways of studying chromatin dynamics and organization, and provide new avenues for understanding the evolutionary role of large-scale GRs. Genomes are not merely a composite of linear DNA sequences, but are also highly organized and regulated inside the nucleus. In an attempt to address the apparent conundrum posed by the contrasting outcomes of genomic organization, and to further develop a holistic appreciation of genome complexity, we introduce a novel, multidisciplinary hypothesis for the study of genome evolution: the *Integrative Breakage Model*. We propose that both the composition of DNA sequences and chromatin conformation are key elements necessary to understand how, and at what point during the cell cycle, new chromosomal forms originate and are subsequently passed to the offspring. This highlights the central role played by chromatin structure in determining the evolution of genomic architecture.

Origin and genomic distribution of evolutionary breakpoints

xEvolutionary biologists have long sought to understand the mechanisms underlying GR across the Tree of Life. This has led to a substantial literature that has focused on tracking GR in vertebrates (with an evolutionary history dating back to 450 million years ago, mya), amniotes (310 mya), and tetrapods (360 mya) [2–7]. It is apparent that each phylogenetic lineage has followed an independent pattern. Mammals, for example, show significant chromosomal number variation (from $2n = 6$ to $2n = 106$), while birds have a very stable karyotype ($2n = 80$) (reviewed in [6]), raising questions as to what drives these apparently independent GR events. The *Random Breakage Model* by Nadeau and Taylor [8] was the first attempt to explain how genomes evolve using human and mouse linkage maps. Their hypothesis relied on two main assumptions: first, many (and large) chromosomal segments are expected to be conserved among related species (so-called Homologous Synteny Blocks, HSBs) and, secondly, that GRs would occur uniformly across genomes and independently of each other (i.e. any genomic region has an equal probability to reorganize). Reanalysis of the increasingly large numbers of orthologous genes identified in these genomes confirmed initial observations [9]. However, following the publication of the first draft of the human and mouse genomes, the second of Nadeau and Taylor's postulates was questioned by the Fragile Breakage Model [10] which initially showed that some breakpoints occurred more than once, suggesting their reuse since divergence from the human–mouse common ancestor [11]. Importantly, Pevzner and Tesler [11] did not provide specific locations for the genomic regions that have been reused, but relied rather on a mathematical model that probed for the existence of such regions. Empirical evidence of breakpoint reuse from cross-species fluorescence in situ hybridization experiments (known as Zoo-FISH or chromosome painting) comprising data from more >100 mammalian species [12–15], as well as from whole genome comparisons [6, 16–18], soon followed. These comparative studies clearly demonstrated that the sites at which structural changes occur (evolutionary breakpoint regions, EBRs), are not evenly distributed, but rather cluster in so-called "hotspots" of GR and, in agreement with [10], that these regions were reused. It is important to emphasize that in a strictly phylogenetic context, the term "breakpoint reuse" accounts for the occurrence of the same breakpoint in two species that do not share a recent common ancestor (i.e. pointing to a polyphyletic origin for these breakpoints [16–19]).

The idea that some chromosomal regions have been reused during mammalian evolution led to investigations to determine whether DNA configuration (and/or its composition) was responsible for genomic instability. Are these regions labile because of their DNA sequence and/or structural chromatin conformation, or do they merely represent regions where selection against breakpoints is reduced or absent?

EBRs are prone to break: The role of repetitive sequences

The question of whether or not a particular DNA sequence drives GR has enjoyed considerable attention. Mammalian breakpoint regions are rich in repetitive elements, including segmental duplications (SDs) [20–22], tandem repeats (TRs) [17, 19, 23], transposable elements (TEs) [19, 24–26]. This is particularly well illustrated by human–chimpanzee comparisons (Box 1) where six of the nine pericentric inversions have EBRs enriched with SDs and repetitive elements (TRs and TEs) [21], suggesting a common mechanistic role in their formation. Since SDs and TEs are regions with a high degree of homology, Non-Allellic Homologous Recombination (NAHR) most likely promotes chromosomal rearrangements by using SDs and/or TEs as templates [27, 28]. In fact, breakpoints enriched for TEs have been reported not only in mammals [21, 25, 29, 30] but also in *Drosophila* [31], suggesting involvement across diverse lineages. TRs, on the other hand, are an important source of DNA variation and mutation [32]. They are considered important facilitators of GR as a result of their capacity to form a variety of secondary DNA structures that include hairpins and bipartite triplexes [33]. The instability associated with the presence of TRs is thought to result from DNA polymerase slippage during DNA replication and recombination during meiosis [34–36]. It seems probable that just as TRs are affected by deletions and expansions in some well-known human diseases [35], so too are they implicated in the formation of GRs during genome evolution [17, 19], since they may fold into secondary DNA structures and induce DSBs [37, 38]. In some aspects repetitive elements represent a complex mix of repeat types—SDs are related to TEs and these in turn can generate microsatellites. SDs, for example, have transposable elements at their terminal regions [39]. In fact, a high proportion of *Alu* elements (a class of TE present only in primates) have been detected at the ends of SDs in the human genome, suggesting that they were generated by *Alu* mispairing followed by homologous recombination [39]. On the other hand,

Box 1

Chromosomal rearrangements in Great Apes

Work by Yunis and Prakash [108] is considered foundational in defining the chromosomal rearrangements that distinguish human and chimpanzee karyotypes since their divergence from a common ancestor ~4–6 mya [109]. Differences include nine inversions and one fusion of two ancestral hominoid chromosomes to produce the modern form of human chromosome 2 (see Fig. below). The inversions affecting human chromosome 1 and 18 have been fixed in the human lineage, whereas the remainder (on orthologs 4, 5, 9, 12, 15, 16, and 17) are autapomorphies that distinguish the chimpanzee lineage [21, 64]. Chromosome 10 underwent one inversion that is fixed in humans and chimpanzees, and a new inversion that was subsequently fixed in gorilla. Chromosome 12 is maintained in the ancestral form in humans and orangutans but underwent an inversion that has been fixed in chimpanzee and gorilla.

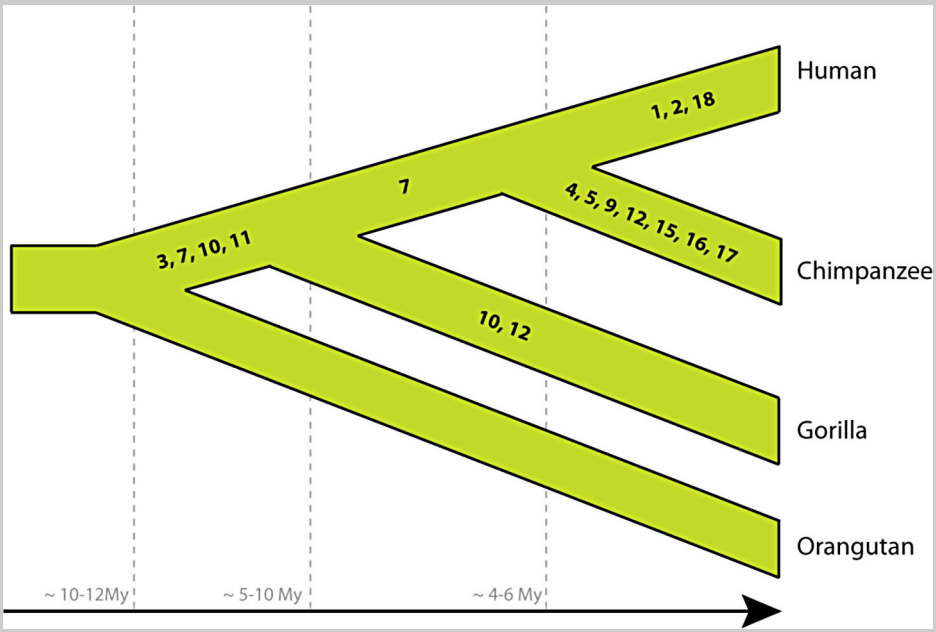

Figure I: Evolutionary history of great apes chromosomes [64]. Phylogeny and divergence times are based on previous studies [109].

Box 2

The combining role of simple repeats and TEs

TEs and microsatellites can, in combination, generate GR. For example, each new copy of an *Alu* element provides two potential sources of microsatellites, the first being the linker region in the middle of the element and the second the 3′ (dA)-rich tail [110]. The role of repeats and TEs was recently examined by Farré and collaborators [19], who analyzed the repeat content in primate EBRs and found that AT-rich motifs accounted for more than 30% of all TRs detected in great apes and rhesus macaque genomes, with the AAAT motif being the most abundant. Interestingly, this motif can form single-stranded coils, thus favoring chromatin instability [37]. Moreover, these authors have found that only primate specific EBRs are enriched for *Alu* elements and that only the AAAT motif was significantly associated with *Alu* elements in EBRs suggesting there was a burst of *Alu* elements in the primate common ancestor at ~40 mya [39]. Given the AAAT motif is similar to the canonical insertion motif 5′-TTAAA-3′ [111] and that *Alu* elements are capable of insertion at target sites that are slightly different (although always AT-rich [112]), the AAAT motif could be site-specific for *Alu* insertion in primate EBRs. This is consistent with observations by Kvikstad and Makova [113] who determined that young *Alu* families were located in AT-rich regions in the human genome. These observations have been extended to other mammalian species including certartiodactyls [30], where lineage-specific EBRs are enriched for LTR endogenous retrovirus 1 (LTR-ERV1), satellite repeats, and tRNAGlu-derived SINEs strongly supporting the hypothesis that active TEs could promote lineage-specific rearrangements [18, 19].

given the high abundance of TEs in mammalian genomes (they make up ~50% of the human genome, [40]), and because they contain homopolymeric tracts, TEs can generate microsatellites themselves [19] (Box 2).

In summary, analysis of a wide range of eukaryotic genomes such as mammals, *Drosophila* [31, 41], *Anopheles* [42], plants [43], and even bacterial genomes [44] provides compelling evidence that EBRs are linked to the presence of repetitive elements. The evidence also suggests that DNA sequence composition plays an important role in the genome-wide distribution of EBRs in eukaryotes. However, given the diversity of repetitive elements in EBRs, it is likely that sequence composition is not alone in influencing genome instability.

Are EBRs regions of low purifying selection?

Mongin and collaborators [45] proposed that long-range transcriptional regulation strongly influences the fixation of chromosome breaks. In many regions, the fitness costs of altering the spatial associations between long-range regulatory regions and their target genes may be so high as to forbid rearrangement. As a consequence one may ask whether EBRs could be considered regions where selection against breakpoints is reduced, or even absent? In addition to the models discussed above, the Intergenic Breakage Model [1] suggests that selection prevents breaks occurring within genes and regulatory regions upstream from genes. The model holds that DSBs (the origin of EBRs) are not located at ''preferred'' sites in the genome. Instead they appear to be random but only in the sense that those that do not disrupt essential genes and/or gene expression actually become fixed. Cells carrying disrupted genes would incur negative selection, and the responsible rearrangement would therefore not be passed to offspring. Consequently EBRs are not considered physically unstable because of their sequence composition, but rather because they develop in areas where selection against breakpoints is minimal.

Initial reports indicated that only DSBs occurring within intergenic regions escape purifying selection and become fixed in the population [46, 47]. Several lines of evidence that show that EBRs in gene dense regions are precisely located between genes support this view (see, for example, [48, 49])—as does the observation that the ~4% of the human genome under selective constraints consists largely of coding regions, introns, and intergenic regions [50]. This effect is most pronounced in promoters of genes that govern development and basic cellular functions. It is somewhat more relaxed in promoters of genes linked to the immune system, reproduction, and perception [50]. Interestingly, studies defining conserved ancestral microsyntenic gene pairs between all sequenced bilaterian animals [51] show that genes from these regions have maintained their genomic linkage in evolution due to gene expression co-regulation, or because they comprise the genomic regulatory landscapes of key transcriptional regulators and developmental genes [52]. This explains why breakpoints in these ancient cis-genomic regulatory blocks might be under purifying selection.

But how universal is this pattern? Certain changes in gene expression could reflect a selective advantage, but empirical data are weak. On the one hand, it has been reported that EBRs co-localize with genes related to the immune system [18, 30, 53], suggesting the connection between EBRs and the development of new adaptive characters may be specific to mammalian lineages. On the other hand, studies in *Drosophila* have shown both changes in gene expression levels resulting from GR [54] and a high number of gene alterations at breakpoints in *D. mojavensis* [55], all with putative adaptive consequences. In reality, however, data on altered gene expression resulting from EBR activity across entire genomes are limited and focused almost exclusively on yeast strains [56], or human and chimpanzee [21, 57]. More confidence in these conclusions will result from greater species representation [58].

Importantly, repetitive elements, and not only genes per se, may be subjected to the action of selection. EBRs can be located in genomic regions rich in SDs (see above), and SDs have, in turn, been associated with lineage-specific copy number variants (CNVs) [59] that are affected by positive selection in coding regions [60]. Changes in the copy number of these highly repetitive regions could play a role in the evolution of adaptive characters specific for each lineage, and this may be reflected by increased expression of genes located in CNVs, or through the regulation of the expression of nearby genes [61, 62]. In a similar vein, 11% of the human gene regulatory sequences conserved among 28 mammalian species were found to be co-opted from ancestral insertions of transposable elements [50, 63], indicating a possible role for these elements in gene expression differences across lineages.

The genealogy of genome reorganization: Descent with modification

As with any evolutionary change of state, GRs must originate in the germ line either before (proliferating oogonia), or during the meiotic division (spermatocytes and oocytes). In such cases GRs can reduce gene flow and potentially contribute to speciation by the suppression of recombination in the reorganized regions between chromosomally different, but contiguous populations [64–67]. However, few empirical data are available that illustrate the mechanisms by which evolutionary regions affect recombination and vice versa. Meiotic recombination in EBRs has been studied in *Drosophila* [68], birds [69], human [18, 64], and chimpanzee [64], all of which show that EBRs are regions with low recombination. The low levels of recombination could lead to a high divergence and fixation of new mutations in these regions which, combined with the presence of genes related to the species-specific phenotypes and biology in these genomic regions, reinforce the adaptive value of EBRs. On the other hand, EBRs are initiated by DSBs, as is meiotic recombination [70]. Consequently, the mechanisms responsible of the formation of DSBs during recombination in germ cell lines could similarly determine the mammalian position of EBRs.

It is noteworthy, therefore, that an association between repeated elements, open chromatin conformations and hotspots for recombination (Box 3) has been detected. Recent genome-wide recombination initiation maps in individual

Box 3

Meiotic recombination hotspots

In meiosis, crossover events are not randomly distributed across the genome. They tend to occur preferentially at ''hotspots'', although little is known of how these regions are selected by the cell machinery [114]. Linkage disequilibrium (LD) analysis has identified over 30,000 hotspots in humans [115] and between 15,000 and 20,000 in the mouse [116]. Hotspots are enriched for degenerative repeat motifs (i.e. a degenerate 13-base-pair repeat motif (CCNCCNTNNCCNC) in the case of humans [115]). These are recognized by the PR domain-containing 9 (PRDM9) protein, a meiotic-specific methyltransferase that catalyses histone H3 lysine 4 trimethylation (H3K4me3) [117]. Studies in mice indicate that the epigenetic modifications introduced by PRDM9 permit the recruitment of the recombination initiation machinery during meiosis, and govern the position at which recombination takes place [116]. Significantly in this context, Prdm9 is the only known mammalian speciation-associated gene [118]. Its sequence composition varies among species and also individuals, affecting recombination rates in natural populations [119]. It has been shown that variation in the Prdm9 sequence affects the positioning of DSBs during meiosis, and that the number and sequence of Zn fingers modulates the strength and specificity of DNA binding [71, 116].

human males have shown that DSB meiotic hotspots are enriched at the breakpoints of CNVs that arise via homology-mediated mechanisms [71]. The presence of degenerate repetitive elements in these regions is probably due to the selective advantages that they provide—the rapid evolution of recombination hotspots. This results in the ''recombination hotspot paradox'': the contradiction between the long-term persistence of recombination hotspots and the self-destructive mechanism by which they are generated [72]. This paradox could explain why EBRs are regions of low recombination—they may have been hotspots in the ancestor which, following GR, became regions of low recombination. As a consequence, hotspots may be considered ''birth and death events,'' fully consistent with Alekseyev and Pevzner's Turnover Fragile Breakage Model (TFBM) [73]. The TFBM argues that fragile regions of the genome are transient states – that is, few regions in a genome are fragile at any given time. In other words, fragile regions blink on and off at different times during the evolution of different lineages (i.e. they are not fixed genomic features). This model explains the different number of EBRs detected in different phylogenetic branches (see Box 1) enforcing the notion that EBRs are no longer active, and exist only as relics of rearrangements that occurred in the evolutionary past. Confirmation is likely to depend on recombination maps for a broader suite of animal taxa that will permit determination of lineage-specific recombination rates.

However, as stated previously, genomes are not simply complexes of linear DNA sequences. The chromatin organization of the nucleus changes during the cell cycle promoting accessibility of the DNA to the replication and repair machinery. Additionally, compartmentalization and the spatial organization of genes and regulatory regions are essential for gene expression. Therefore, studying processes that promote DNA accessibility will advance our understanding of the influence of genome instability on large-scale rearrangements.

Chromatin conformation: Facilitator of genome reorganization?

Genomes are chromatin structures, the regulation of which depends on several superimposed layers of organization that include: (i) the chemical modification of DNA, (ii) the presence of nucleosomes that wrap the DNA around four core histones (H2A, H2B, H3, and H4), (iii) the high-order organization of chromatin compartments inside the nucleus, and (iv) gene expression inside chromatin compartments during the cell cycle and during cell differentiation. Points (i) and (ii) constitute what is known as the epigenome and (iii) and (iv) 3D and 4D chromatin architecture, respectively. The way in which these different levels of chromatin organization interact in vivo during the cell cycle, and in early development and/or cell differentiation, determines organizational plasticity, an area of research that is largely unexplored, especially so in an evolutionary context.

The evolutionary epigenomic landscape: The influence of genome reorganization

The epigenome includes all chemical changes in the DNA and histone proteins of a genome that influence gene expression and genome organization [74, 75]. These include DNA methylation at cytosine residues, as well as histone methylation and acetylation [76]. The general picture to emerge from studies on human cells is that certain histone modifications [acetylation (ac) or methylation (me)] of residues (H3K9ac, H3K27ac, and H3K4me3) are markers of open chromatin conformation and regions of active transcription (reviewed in [77]). Other modifications, including H3K27me3 or H3K9me3, are related to ''closed'' chromatin conformations, and correspond to regions where gene transcription is inactive. Most studies have been performed on somatic cells where ''open'' chromatin conformations are rich in genes and CpG islands [78], and are characterized by low levels of DNA methylation [63, 64]. Recently the ENCODE consortium [79] provided an exhaustive analysis of the chromatin elements that regulate gene expression

and genome organization. This makes it possible to detect the presence of potential ''open/closed'' configurations in the human genome. Among these elements are DNase I hypersensitive sites (DHSs), considered markers of regulatory DNA and nuclease binding sites, and therefore markers of accessible chromatin (i.e. nucleosome-free regions). ENCODE lists ~2.9 million DHSs in 125 diverse human cell lines and tissues, but only 3,692 were conserved in all samples. Interestingly, 5% of these DHSs are located within 2.5 Kbp of the genomic sequence (excluding transcriptional start sites), some of which are in TEs [79]. Moreover, the authors found a negative correlation between DHSs and methylation, suggesting that accessible chromatin is usually less methylated.

Significantly, however, little is known of the evolutionary relationship between the epigenome and GRs. Lemaitre and collaborators [48] were the first to establish a link between genome organization and the distribution of mammalian EBRs. They observed that evolutionary rearrangements tend to occur in regions of high transcriptional activity because of the ''open'' conformation (decondensed chromatin state) of these regions [48]. Likewise, high breakpoint density was associated with under-methylated regions of the genome, and hence the accessible state of the chromatin. This is further facilitated by transposons that integrate into nucleosome-free regions in vivo [80, 81], promoting genome instability through homology-mediated mechanisms. In similar fashion, Carbone et al. [82] recently found that EBRs in gibbons (a primate species with extensive GRs [83, 84]) were especially rich in *Alu* elements. These elements are remarkably hypomethylated compared to human *Alu* [24], suggesting that the epigenome plays a role in the distribution of large-scale genomic changes by determining the accessibility/susceptibility of chromatin to alteration.

New views on the nucleome and its role in genome organisation

Another important factor that can influence the distribution of EBRs across genomes is the three-dimensional (3D) structure of the nucleus, also known as nucleome [85], an aspect that has often been overlooked in evolutionary studies. The genome is organized into discrete, 3D chromosomal territories or domains [86], which interact among themselves inside the nucleus. Such organization is important for understanding biological function and gene expression/regulation and how these processes govern genome reshuffling. For GR to result, two DSBs have to occur in the germ line followed by misrepair. For instance, in a translocation two DSBs will occur and the cell machinery will repair them using a non-homologous chromosome as a template. This is only possible if the regions involved in the rearrangement are in close proximity to each other.

Initial studies used high-definition microscopy to show that the organization of the genome inside the nucleus is non-random. Gene-rich chromosomes and active euchromatin tend to reside in the inner portion of nuclei, while gene-poor regions and genetically inert heterochromatin are located at the nuclear periphery [86, 87]. This peculiar interphase positioning is evolutionarily conserved in mammals [88, 89] and facilitates the physical localization of co-regulated genomic domains in the nucleus, resulting in transcription factories [90]. The more detailed analysis of genome-wide physical interactions afforded by recent chromosome conformation capture techniques (3C) and derivatives (4C, 5C, and Hi-C) [91], makes it possible to study chromosomal interactions in vivo by analyzing the average frequency at which genomic loci are physically associated inside the nucleus within cell populations [91]. This allows the structural properties and the spatial organization of chromosomes to be examined by high-throughput technologies, providing new insights into the regulation of gene expression, DNA replication and repair, and recombination. This has revealed different levels of hierarchical genome organization: (i) chromosomal territories, (ii) ''open'' (termed ''A'')/''closed'' (termed ''B'') compartments inside chromosomal territories, (iii) topological associated domains (TADs), and (iv) looping interactions [92–94] (Fig. 1). Chromosomal territories contain A/B compartments which are composed of TADs—discrete chromatin domains (800 kb medium size) that contain genomic loci that have a high tendency to interact among themselves (more so than with loci outside the TAD). The fact that TADs are present in taxa as diverse as human, mouse, and *Drosophila*, and are stable across different cell types and developmental stages [95–99], reinforces their importance as regulators of the genomic landscape. Moreover, TADs boundaries are enriched for genomic elements, such as the CCCTC-binding factor (CTCF), suggesting the boundaries act as insulators (i.e. elements that block the interaction between enhancers and promoters) [95, 100].

An integrative model for the origin of evolutionary breakpoints

One may be forgiven for thinking that there is little commonality among the models presented. On the one hand, studies have shown that genomic regions are prone to breakage due to the presence of repetitive sequences [17–19, 25]; on the other, initial comparative results [19, 24, 48] suggested that certain properties of DNA sequences (such as repetitive elements together with the epigenetic state) could promote open DNA chromatin configurations during the cell cycle that lead to GR. Conversely, purifying selection has been mooted as the primary driver of EBR distribution in a variety of genomes [1, 46, 48, 101]. We are of the view, however, that these models are not mutually exclusive, and that they can be accommodated in a new *Integrative Breakage Model* of GR that incorporates (i) DNA sequence composition, (ii) the nucleome, and (iii) the effect on gene expression as key elements in determining the genomic distribution of evolutionary breakpoints (Figs. 1 and 2). There is a substantial literature that implicates sequence composition as a cause of genomic plasticity but, importantly, only DSBs occurring in regions not affecting DNA secondary structure, or that do not modify the expression of genes related to development or basic cellular maintenance (i.e. housekeeping genes), have the potential to become fixed and are therefore of evolutionary significance.

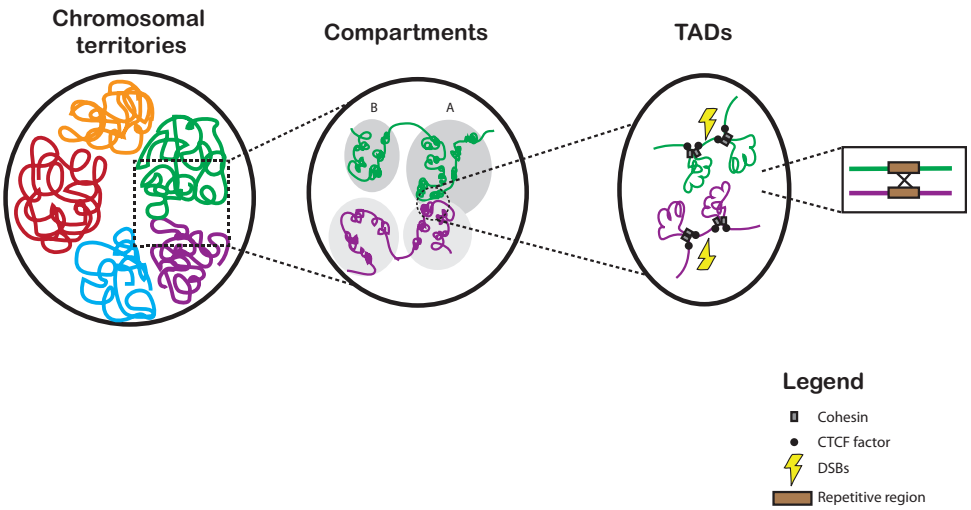

Figure 1. Chromatin conformation in eukaryotic cells and its involvement in genome reshuffling. Chromosomes occupy distinct territories in the interphase nucleus, which are compartmentalized into open (''A'') or closed (''B'') chromatic regions (see text for further details). Both compartments contain Topological Associated Domains (TADs) delimited by cohesins and CTCF factors. Double-strand breaks (DSBs) will occur between two TADs and can recombine and resolve into a genome rearrangement by means of Non-Allelic Homologous Recombination (NAHR) using repetitive regions as templates.

Additionally, our model explains the presence of multispecies HSBs (i.e. genomic blocks conserved in several species) that are enriched for gene networks that control embryonic and tissue development (basic functions necessary for the organisms) and why genes that facilitate adaptive responses tend to localize in EBRs [18, 102].

In fact, 3C high-throughput methodologies have made it possible to test whether genomic regions that tend to break and reorganize interact inside the nucleus. Using the Hi-C data from the human genome [103], Veron and colleagues [104] showed that orthologous sequences that were distantly located loci in the human genome, but closely

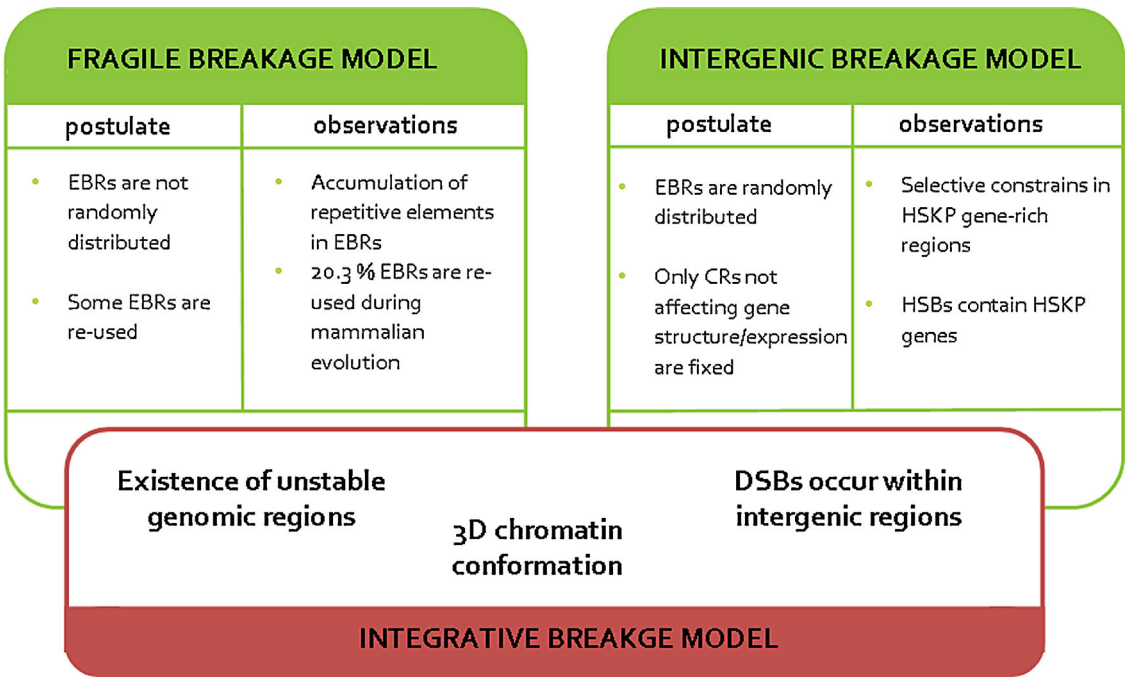

Figure 2. The Integrative Breakage Model of genome evolution. EBRs, evolutionary breakpoint regions; HSBs, homologous synteny blocks; TR, tandem repeats; SDs, segmental duplications; CRs, chromosomal reorganizations; HSKP, house-keeping genes; 3D, three-dimensional; DSBs, double-strand breaks.

located in the mouse genome (and, therefore, affected by GRs), were closer in human interphase nuclei. In similar fashion, mouse cells reveal that regions involved in translocations are found in close proximity in interphase nuclei [105] suggesting that while breakpoint frequency might be similar across all chromosomes, the frequency of translocations is probably related to the relative proximity of chromosomal breakpoints. These data collectively suggest an important role for both the epigenetic state and the nucleome in the distribution of EBRs. The probability of interchromosomal reorganization occurring is complex: DSBs develop, and the close proximity of "open" chromatin conformations, such as TADs and their boundaries, facilitate exchanges (Fig. 1). If these DSBs are produced at early stages of embryonic development or in the germ line they can result in structural rearrangements that are transmitted as potentially new chromosomal variants that are associated with some selective advantage, that may even facilitate speciation.

However, for a reorganization to spread among individuals within a population it is necessary to invoke some form of selection. The *Integrative Breakage Model* is also consistent with the concept of a "functional neighbourhood" [106] whereby genes are arranged in regions according to functional characteristics. The presence of TADs and their boundaries in taxa as diverse as human, mouse, and *Drosophila* may be a result of "cis" regulatory interactions that are likely to be maintained by purifying selection. This is supported by the presence of ectopic intra-chromosomal contacts and gene expression deregulation when boundaries between consecutive TADs are deleted experimentally [96]. These neighborhoods have a high-level of gene co-expression and they are consistently distributed across closely related species. The function of a cluster of genes would be constrained in these neighborhoods, but this would not influence the individual genes themselves. Put differently, regions with functional equivalency are formed by different (non-orthologous) genes in different species. Remarkably, these neighborhoods are enriched in EBRs showing the likely effects of selection on the transcription of blocks of functionally related genes. Therefore, if a chromosomal rearrangement breaks a functional neighborhood, selection will favor further rearrangements and the reconstruction of a new neighborhood with a similar function. This underscores the highly dynamic nature of the genomic landscape.

Our model of a dynamic genomic landscape characterized by fluctuations in SDs, CNVs, and TE insertion sites is also compatible with the Turnover Fragile Breakage Model (TFBM) [73]. We propose that this "transient state" is related to the changing genomic landscape (i.e. variability in DNA repeat content, chromatin conformation, and distribution) and that this is lineage specific. Moreover, these are linked to the evolutionary conservation of TADs and their genomic boundaries across taxa. Studies on the potential role of CTCF and other zinc finger proteins with insulator properties will provide the clues on how chromatin looping activities are regulated in the cell cycle and subjected to selective constrains. In fact, in mammals CTCF binding sites are composed of a 33/34 bp motif with a two-part profile that is hierarchically conserved across species and associated with repeat element expansions [107].

Conclusions and outlook: The promise of functional genomics

We show that the genomic distribution of mammalian EBRs is multifactorial and depends on repetitive elements, functional constrains, and the nucleome, as well as meiotic recombination. We propose that genomic regions involved in evolutionary reshuffling (i) interact physically inside the nucleus during the formation of the germ line, (ii) they present open chromatin DNA configurations and epigenetic features that could promote DNA accessibility and therefore genomic instability, and (iii) only those reorganizations that do not disturb essential genes and/or gene expression will likely be fixed within populations. These data suggest several potentially useful avenues of further research. For example, the importance of open chromatin regions for the occurrence of GRs could be tested through functional genomics since heritable rearrangements would be expected to occur in those regions that are specifically accessible in the germ line and/or early totipotent developmental stages. We foresee that high-throughput multidisciplinary studies of chromatin interactions, epigenetic signatures, and functional genomic characteristics of eukaryotic genomes will drive future developments in the field, and that these hold promise for improving our understanding of the mechanics and evolution of genomic diversity.

Acknowledgments

We thank the editor and two anonymous referees for their insightful suggestions on improving the manuscript. MF is supported by a BBSRC grant (BB/J010170/1) and TJR by the South African National Research Foundation. ARH's laboratory is supported by the Spanish Ministerio de Ciencia y Tecnologia (CGL-2010–20170).

The authors have declared no conflicts of interest.

References

1. **Peng Q**, **Pevzner PA**, **Tesler G.** 2006. The fragile breakage versus random breakage models of chromosome evolution. *PLoS Comput Biol* **2**: e14.
2. **Postlethwait JH**, **Woods IG**, **Ngo-Hazelett P**, **Yan YL**, et al. 2000. Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res* **10**: 1890–902.
3. **Kohn M**, **Hogel J**, **Vogel W**, **Minich P**, et al. 2006. Reconstruction of a 450-My-old ancestral vertebrate protokaryotype. *Trends Genet* **22**: 203–10.
4. **Nakatani Y**, **Takeda H**, **Kohara Y**, **Morishita S.** 2007. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome Res* **17**: 1254–65.
5. **Uno Y**, **Nishida C**, **Tarui H**, **Ishishita S**, et al. 2012. Inference of the protokaryotypes of amniotes and tetrapods and the evolutionary processes of microchromosomes from comparative gene mapping. *PLoS ONE* **7**: 53027.
6. **Ruiz-Herrera A**, **Farré M**, **Robinson TJ.** 2012. Molecular cytogenetic and genomic insights into chromosomal evolution. *Heredity* **108**: 28–36.
7. **Amemiya CT**, **Alfoldi J**, **Lee AP**, **Fan S**, et al. 2013. The African coelacanth genome provides insights into tetrapod evolution. *Nature* **496**: 311–6.
8. **Nadeau JH**, **Taylor BA.** 1984. Lengths of chromosomal segments conserved since divergence of man and mouse. *Proc Natl Acad Sci USA* **81**: 814–8.
9. **Nadeau JH**, **Sankoff D.** 1998. Counting on comparative maps. *Trends Genet* **14**: 495–501.

10. **Pevzner P**, **Tesler G.** 2003. Genome rearrangements in mammalian evolution: lessons from human and mouse genomes. *Genome Res* **13**: 37–45.
11. **Pevzner P**, **Tesler G.** 2003. Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. *Proc Natl Acad Sci USA* **100**: 7672–7.
12. **Froenicke L.** 2005. Origins of primate chromosomes—as delineated by Zoo-FISH and alignments of human and mouse draft genome sequences. *Cytogenet Genome Res* **108**: 122–38.
13. **Ruiz-Herrera A**, **Garcia F**, **Mora L**, **Egozcue J**, et al**.** 2005. Evolutionary conserved chromosomal segments in the human karyotype are bounded by unstable chromosome bands. *Cytogenet Genome Res* **108**: 161–74.
14. **Robinson TJ**, **Ruiz-Herrera A**, **Froenicke L.** 2006. Dissecting the mammalian genome-new insights into chromosomal evolution. *Trends Genet* **22**: 297–301.
15. **Froenicke L**, **Caldes MG**, **Graphodatsky A**, **Muller S**, et al**.** 2006. Are molecular cytogenetics and bioinformatics suggesting diverging models of ancestral mammalian genomes. *Genome Res* **16**: 306–10.
16. **Murphy WJ**, **Larkin DM**, **Everts-van der Wind A**, **Bourque G**, et al**.** 2005. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science* **309**: 613–7.
17. **Ruiz-Herrera A**, **Castresana J**, **Robinson TJ.** 2006. Is mammalian chromosomal evolution driven by regions of genome fragility. *Genome Biol* **7**: R115.
18. **Larkin DM**, **Pape G**, **Donthu R**, **Auvil L**, et al**.** 2009. Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome Res* **19**: 770–7.
19. **Farré M**, **Bosch M**, **López-Giráldez F**, **Ponsà M**, et al**.** 2011. Assessing the role of tandem repeats in shaping the genomic architecture of great apes. *PLoS ONE* **6**: e27239.
20. **Bailey JA**, **Eichler EE.** 2006. Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat Rev Genet* **7**: 552–64.
21. **Kehrer-Sawatzki H**, **Cooper DN.** 2008. Molecular mechanisms of chromosomal rearrangement during primate evolution. *Chromosome Res* **16**: 41–56.
22. **Zhao H**, **Bourque G.** 2009. Recovering genome rearrangements in the mammalian phylogeny. *Genome Res* **19**: 934–42.
23. **Kehrer-Sawatzki H**, **Sandig C**, **Chuzhanova N**, **Goidts V**, et al**.** 2005. Breakpoint analysis of the pericentric inversion distinguishing human chromosome 4 from the homologous chromosome in the chimpanzee (*Pan troglodytes*). *Hum Mutat* **25**: 45–55.
24. **Carbone L**, **Harris RA**, **Vessere GM**, **Mootnick AR**, et al**.** 2009. Evolutionary breakpoints in the gibbon suggest association between cytosine methylation and karyotype evolution. *PLoS Genet* **5**: e1000538.
25. **Longo MS**, **Carone DM**, **Program NCS**, **Green ED**, et al**.** 2009. Distinct retroelement classes define evolutionary breakpoints demarcating sites of evolutionary novelty. *BMC Genomics* **10**: 334.
26. **Bourque G.** 2009. Transposable elements in gene regulation and in the evolution of vertebrate genomes. *Curr Opin Genet Dev* **19**: 607–12.
27. **Bailey JA**, **Baertsch R**, **Kent WJ**, **Haussler D**, et al**.** 2004. Hotspots of mammalian chromosomal evolution. *Genome Biol* **5**: R23.
28. **Cordaux R**, **Batzer MA.** 2009. The impact of retrotransposons on human genome evolution. *Nat Rev Genet* **10**: 691–703.
29. **Lee J**, **Han K**, **Meyer TJ**, **Kim HS**, et al**.** 2008. Chromosomal inversions between human and chimpanzee lineages caused by retrotransposons. *PloS ONE* **3**: e4047.
30. **Groenen MA**, **Archibald AL**, **Uenishi H**, **Tuggle CK**, et al**.** 2012. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**: 393–8.
31. **Caceres M**, **Ranz JM**, **Barbadilla A**, **Long M**, et al**.** 1999. Generation of a widespread Drosophila inversion by a transposable element. *Science* **285**: 415–8.
32. **Armour JA.** 2006. Tandemly repeated DNA: why should anyone care. *Mutat Res* **598**: 6–14.
33. **Catasti P**, **Chen X**, **Mariappan SV**, **Bradbury EM**, et al**.** 1999. DNA repeats in the human genome. *Genetica* **106**: 15–36.
34. **Usdin K**, **Grabczyk E.** 2000. DNA repeat expansions and human disease. *Cell Mol Life Sci* **57**: 914–31.
35. **Shaw CJ**, **Lupski JR.** 2004. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Hum Mol Genet* 13 Spec No **1**: R57–64.
36. **Kelkar YD**, **Tyekucheva S**, **Chiaromonte F**, **Makova KD.** 2008. The genome-wide determinants of human and chimpanzee microsatellite evolution. *Genome Res* **18**: 30–38.
37. **Bacolla A**, **Larson JE**, **Collins JR**, **Li J**, et al**.** 2008. Abundance and length of simple repeats in vertebrate genomes are determined by their structural properties. *Genome Res* **18**: 1545–53.
38. **Zhao J**, **Bacolla A**, **Wang G**, **Vasquez KM.** 2010. Non-B DNA structure-induced genetic instability and evolution. *Cell Mol Life Sci* **67**: 43–62.
39. **Bailey JA**, **Eichler EE.** 2003. Genome-wide detection and analysis of recent segmental duplications within mammalian organisms. *Cold Spring Harb Symp Quant Biol* **68**: 115–24.
40. **Lander ES**, **Linton LM**, **Birren B**, **Nusbaum C**, et al**.** 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
41. **González J**, **Casals F**, **Ruiz A.** 2007. Testing chromosomal phylogenies and inversion breakpoint reuse in Drosophila. *Genetics* **175**: 167–77.
42. **Lobo NF**, **Sangaré DM**, **Regier AA**, **Reidenbach KR**, et al**.** 2010. Breakpoint structure of the *Anopheles gambiae* 2Rb chromosomal inversion. *Malar J* **9**: 293.
43. **dela Paz JS**, **Stronghill PE**, **Douglas SJ**, **Saravia S**, et al**.** 2012. Chromosome fragile sites in Arabidopsis harbor matrix attachment regions that may be associated with ancestral chromosome rearrangement events. *PLoS Genet* **8**: e1003136.
44. **Darling AE**, **Miklós I**, **Ragan MA.** 2008. Dynamics of genome rearrangement in bacterial populations. *PLoS Genet* **4**: e1000128.
45. **Mongin E**, **Dewar K**, **Blanchette M.** 2009. Long-range regulation is a major driving force in maintaining genome integrity. *BMC Evol Biol* **9**: 203.
46. **Becker TS**, **Lenhard B.** 2007. The random versus fragile breakage models of chromosome evolution: a matter of resolution. *Mol Genet Genom* **278**: 487–91.
47. **Kikuta H**, **Laplante M**, **Navratilova P**, **Komisarczuk AZ**, et al**.** 2007. Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res* **17**: 545–55.
48. **Lemaitre C**, **Zaghloul L**, **Sagot MF**, **Gautier C**, et al**.** 2009. Analysis of fine-scale mammalian evolutionary breakpoints provides new insight into their relation to genome organisation. *BMC Genomics* **10**: 335.
49. **Kemkemer C**, **Kohn M**, **Cooper DN**, **Froenicke L**, et al**.** 2009. Gene synteny comparisons between different vertebrates provide new insights into breakage and fusion events during mammalian karyotype evolution. *BMC Evol Biol* **9**: 84.
50. **Lindblad-Toh K**, **Garber M**, **Zuk O**, **Lin M**, et al**.** 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**: 476–82.
51. **Simakov O**, **Marletaz F**, **Cho S-J**, **Edsinger-Gonzales E**, et al**.** 2013. Insights into bilaterian evolution from three spiralian genomes. *Nature* **493**: 526–31.
52. **Irimia M**, **Maeso I**, **Roy SW**, **Fraser HB.** 2013. Ancient cis-regulatory constraints and the evolution of genome architecture. *Trends Genet* **29**: 521–8.
53. **Ullastres A**, **Farré M**, **Capilla L**, **Ruiz-Herrera A.** 2014. Unraveling the effect of genomic structural changes in the rhesus macaque - implications for the adaptive role of inversions. *BMC Genomics* **15**: 530.
54. **Puig M**, **Caceres M**, **Ruiz A.** 2004. Silencing of a gene adjacent to the breakpoint of a widespread Drosophila inversion by a transposon-induced antisense RNA. *Proc Natl Acad Sci USA* **101**: 9013–8.
55. **Guillen Y**, **Ruiz A.** 2012. Gene alterations at Drosophila inversion breakpoints provide prima facie evidence for natural selection as an explanation for rapid chromosomal evolution. *BMC Genomics* **13**: 53.
56. **Avelar AT**, **Perfeito L**, **Gordo I**, **Ferreira MG.** 2013. Genome architecture is a selectable trait that can be maintained by antagonistic pleiotropy. *Nat Commun* **4**: 2235.
57. **Marquès-Bonet T**, **Cáceres M**, **Bertranpetit J**, **Preuss TM**, et al**.** 2004. Chromosomal rearrangements and the genomic distribution of gene-expression divergence in humans and chimpanzees. *Trends Genet* **20**: 524–9.
58. **Muñoz A**, **Sankoff D.** 2012. Detection of gene expression changes at chromosomal rearrangement breakpoints in evolution. *BMC Bioinformatics* **13**: S6.
59. **Gazave E**, **Darre F**, **Morcillo-Suarez C**, **Petit-Marty N**, et al**.** 2011. Copy number variation analysis in the great apes reveals species-specific patterns of structural variation. *Genome Res* **21**: 1626–39.
60. **Marques-Bonet T**, **Girirajan S**, **Eichler EE.** 2009. The origins and impact of primate segmental duplications. *Trends Genet* **25**: 443–54.
61. **Orozco LD**, **Cokus SJ**, **Ghazalpour A**, **Ingram-Drake L**, et al**.** 2009. Copy number variation influences gene expression and metabolic traits in mice. *Hum Mol Genet* **18**: 4118–29.
62. **Chaignat E**, **Yahya-Graison EA**, **Henrichsen CN**, **Chrast J**, et al**.** 2011. Copy number variation modifies expression time courses. *Genome Res* **21**: 106–13.
63. **Lowe CB**, **Haussler D.** 2012. 29 Mammalian genomes reveal novel exaptations of mobile elements for likely regulatory functions in the human genome. *PLoS ONE* **7**: e43128.

64. **Farre M**, **Micheletti D**, **Ruiz-Herrera A.** 2013. Recombination rates and genomic shuffling in human and chimpanzee-a new twist in the chromosomal speciation theory. *Mol Biol Evol* **30**: 853–64.
65. **Noor MA**, **Grams KL**, **Bertucci LA**, **Reiland J.** 2001. Chromosomal inversions and the reproductive isolation of species. *Proc Natl Acad Sci USA* **98**: 12084–8.
66. **Rieseberg LH.** 2001. Chromosomal rearrangements and speciation. *Trends Ecol Evol* **16**: 351–8.
67. **Faria R**, **Navarro A.** 2010. Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol Evol* **25**: 660–9.
68. **Navarro A**, **Betran E**, **Barbadilla A**, **Ruiz A.** 1997. Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. *Genetics* **146**: 695–709.
69. **Romanov MN**, **Farré M**, **Lithgow PE**, **Fowler KE**, et al**.** 2014. Reconstruction of gross avian genome structure, organization and evolution suggests that the chicken lineage most closely resembles the dinosaur avian ancestor. *BMC Genomics* **15**: 1060.
70. **Keeney S**, **Giroux CN**, **Kleckner N.** 1997. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* **88**: 375–84.
71. **Pratto F**, **Brick K**, **Khil P**, **Smagulova F**, et al**.** 2014. DNA recombination. Recombination initiation maps of individual human genomes. *Science* **346**: 1256442.
72. **Ubeda F**, **Wilkins JF.** 2011. The Red Queen theory of recombination hotspots. *J Evol Biol* **24**: 541–53.
73. **Alekseyev MA**, **Pevzner PA.** 2010. Comparative genomics reveals birth and death of fragile regions in mammalian evolution. *Genome Biol* **11**: R117.
74. **Green CM**, **Almouzni G.** 2002. When repair meets chromatin First in series on chromatin dynamics. *EMBO Rep* **3**: 28–33.
75. **Gontijo AMdMC**, **Green CM**, **Almouzni G.** 2003. Repairing DNA damage in chromatin. *Biochimie* **85**: 1133–47.
76. **Barski A**, **Cuddapah S**, **Cui K**, **Roh T-Y**, et al**.** 2007. High-resolution profiling of histone methylations in the human genome. *Cell* **129**: 823–37.
77. **Campos EI**, **Reinberg D.** 2009. Histones: annotating chromatin. *Annu Rev Genet* **43**: 559–99.
78. **Terrenoire E**, **McRonald F**, **Halsall JA**, **Page P**, et al**.** 2010. Immunostaining of modified histones defines high-level features of the human metaphase epigenome. *Genome Biol* **11**: R110.
79. **Consortium EP**, **Bernstein BE**, **Birney E**, **Dunham I**, et al**.** 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74.
80. **Gangadharan S**, **Mularoni L**, **Fain-Thornton J**, **Wheelan SJ**, et al**.** 2010. DNA transposon Hermes inserts into DNA in nucleosome-free regions in vivo. *Proc Natl Acad Sci USA* **107**: 21966–72.
81. **Buenrostro JD**, **Giresi PG**, **Zaba LC**, **Chang HY**, et al**.** 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin DNA-binding proteins and nucleosome position. *Nat Methods* **10**: 1213–8.
82. **Carbone L**, **Harris RA**, **Gnerre S**, **Veeramah KR**, et al**.** 2014. Gibbon genome and the fast karyotype evolution of small apes. *Nature* **513**: 195–201.
83. **Muller S**, **Wienberg J.** 2001. ''Bar-coding'' primate chromosomes: molecular cytogenetic screening for the ancestral hominoid karyotype. *Hum Genet* **109**: 85–94.
84. **Capozzi O**, **Carbone L**, **Stanyon RR**, **Marra A**, et al**.** 2012. A comprehensive molecular cytogenetic analysis of chromosome rearrangements in gibbons. *Genome Res* **22**: 2520–8.
85. **Pennisi E.** 2015. Genomics. Inching toward the 3D genome. *Science* **347**: 10.
86. **Cremer T**, **Cremer C.** 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* **2**: 292–301.
87. **Cremer T**, **Cremer M.** 2010. Chromosome territories. *Cold Spring Harb Perspect Biol* **2**: a003889.
88. **Tanabe H**, **Müller S**, **Neusser M**, **von Hase J**, et al**.** 2002. Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc Natl Acad Sci USA* **99**: 4424–9.
89. **Mora L**, **Sánchez I**, **Garcia M**, **Ponsà M.** 2006. Chromosome territory positioning of conserved homologous chromosomes in different primate species. *Chromosoma* **115**: 367–75.
90. **Pombo A**, **Branco MR.** 2007. Functional organisation of the genome during interphase. *Curr Opin Genet Dev* **17**: 451–5.
91. **Dekker J**, **Rippe K**, **Dekker M**, **Kleckner N.** 2002. Capturing chromosome conformation. *Science* **295**: 1306–11.
92. **Dekker J**, **Marti-Renom MA**, **Mirny LA.** 2013. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat Rev Genet* **14**: 390–403.
93. **Phillips-Cremins JE.** 2014. Unraveling architecture of the pluripotent genome. *Curr Opin Cell Biol* **28**: 96–104.
94. **Nora EP**, **Dekker J**, **Heard E.** 2013. Segmental folding of chromosomes: a basis for structural and regulatory chromosomal neighborhoods?. *BioEssays* **35**: 818–28.
95. **Dixon JR**, **Selvaraj S**, **Yue F**, **Kim A**, et al**.** 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**: 376–80.
96. **Nora EP**, **Lajoie BR**, **Schulz EG**, **Giorgetti L**, et al**.** 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**: 381–85.
97. **Sexton T**, **Kurukuti S**, **Mitchell JA**, **Umlauf D**, et al**.** 2012. Sensitive detection of chromatin coassociations using enhanced chromosome conformation capture on chip. *Nat Protoc* **7**: 1335–50.
98. **Woltering JM**, **Noordermeer D**, **Leleu M**, **Duboule D.** 2014. Conservation and divergence of regulatory strategies at Hox Loci and the origin of tetrapod digits. *PLoS Biol* **12**: e1001773.
99. **Ghavi-Helm Y**, **Klein FA**, **Pakozdi T**, **Ciglar L**, et al**.** 2014. Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**: 96–100.
100. **Rao SSP**, **Huntley MH**, **Durand NC**, **Stamenova EK**, et al**.** 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**: 1665–80.
101. **Ruiz-Herrera A**, **Farre M**, **Ponsa M**, **Robinson TJ.** 2010. Selection against Robertsonian fusions involving housekeeping genes in the house mouse: integrating data from gene expression arrays and chromosome evolution. *Chromosome Res* **18**: 801–8.
102. **Bovine Genome Sequencing and Analysis Consortium.** 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* **324**: 522–8.
103. **Lieberman-Aiden E**, **van Berkum NL**, **Williams L**, **Imakaev M**, et al**.** 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**: 289–93.
104. **Veron AS**, **Lemaitre C**, **Gautier C**, **Lacroix V**, et al**.** 2011. Close 3D proximity of evolutionary breakpoints argues for the notion of spatial synteny. *BMC Genomics* **12**: 303.
105. **Zhang Y**, **McCord RP**, **Ho YJ**, **Lajoie BR**, et al**.** 2012. Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* **148**: 908–21.
106. **Al-Shahrour F**, **Minguez P**, **Marques-Bonet T**, **Gazave E**, et al**.** 2010. Selection upon genome architecture: conservation of functional neighborhoods with changing genes. *PLoS Comput Biol* **6**: e1000953.
107. **Schmidt D**, **Schwalie PC**, **Wilson MD**, **Ballester B**, et al**.** 2012. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148**: 335–48.
108. **Yunis JJ**, **Prakash O.** 1982. The origin of man: a chromosomal pictorial legacy. *Science* **215**: 1525–30.
109. **Prado-Martinez J**, **Sudmant PH**, **Kidd JM**, **Li H**, et al**.** 2013. Great ape genetic diversity and population history. *Nature* **499**: 471–5.
110. **Shen MR**, **Batzer MA**, **Deininger PL.** 1991. Evolution of the master Alu gene(s). *J Mol Evol* **33**: 311–20.
111. **Jurka J.** 1997. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc Natl Acad Sci USA* **94**: 1872–7.
112. **Levy A**, **Schwartz S**, **Ast G.** 2010. Large-scale discovery of insertion hotspots and preferential integration sites of human transposed elements. *Nucleic Acids Res* **38**: 1515–30.
113. **Kvikstad EM**, **Makova KD.** 2010. The (r) evolution of SINE versus LINE distributions in primate genomes: sex chromosomes are important. *Genome Res* **20**: 600–13.
114. **Jeffreys AJ**, **Kauppi L**, **Neumann R.** 2001. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat Genet* **29**: 217–22.
115. **Myers S**, **Bottolo L**, **Freeman C**, **McVean G**, et al**.** 2005. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**: 321–4.
116. **Brick K**, **Smagulova F**, **Khil P**, **Camerini-Otero RD**, et al**.** 2012. Genetic recombination is directed away from functional genomic elements in mice. *Nature* **485**: 642–5.
117. **Wu H**, **Mathioudakis N**, **Diagouraga B**, **Dong A**, et al**.** 2013. Molecular basis for the regulation of the H3K4 methyltransferase activity of PRDM9. *Cell Rep* **5**: 13–20.
118. **Mihola O**, **Trachtulec Z**, **Vlcek C**, **Schimenti JC**, et al**.** 2009. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* **323**: 373–5.
119. **Capilla L**, **Medarde N**, **Alemany-Schmidt A**, **Oliver-Bonet M**, et al**.** 2014. Genetic recombination variation in wild Robertsonian mice: on the role of chromosomal fusions and Prdm9 allelic background. *Proc Biol Sci* **281**: pii: 20140297.