

Kent Academic Repository

Full text document (pdf)

Citation for published version

Deakin, Janine E. and Potter, Sally and O'Neill, Rachel and Ruiz-Herrera, Aurora and Cioffi, Marcelo B. and Eldridge, Mark D.B. and Fukui, Kichi and Marshall Graves, Jennifer A. and Griffin, Darren and Grutzner, Frank and Kratochvíl, Lukáš and Miura, Ikuo and Rovatsos, Michail and Srikulnath, Kornorn and Wapstra, Erik and Ezaz, Tariq (2019) Chromosomics: Bridging the

DOI

<https://doi.org/10.3390/genes10080627>

Link to record in KAR

<https://kar.kent.ac.uk/76016/>

Document Version

Publisher pdf

Copyright & reuse

Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

Versions of research

The version in the Kent Academic Repository may differ from the final published version.

Users are advised to check <http://kar.kent.ac.uk> for the status of the paper. **Users should always cite the published version of record.**

Enquiries









For any further enquiries regarding the licence status of this document, please contact:

researchsupport@kent.ac.uk

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at <http://kar.kent.ac.uk/contact.html>

Review

Chromosomics: Bridging the Gap between Genomes and Chromosomes

Janine E. Deakin ^{1,*}, Sally Potter ^{2,3,†}, Rachel O'Neill ^{4,†}, Aurora Ruiz-Herrera ^{5,6,†}, Marcelo B. Cioffi ⁷, Mark D.B. Eldridge ³, Kichi Fukui ⁸, Jennifer A. Marshall Graves ^{1,9}, Darren Griffin ¹⁰, Frank Grutzner ¹¹, Lukáš Kratochvíl ¹², Ikuo Miura ¹³, Michail Rovatsos ¹¹, Kornorn Srikulnath ¹⁴, Erik Wapstra ¹⁵ and Tariq Ezaz ^{1,*}

¹ Institute for Applied Ecology, University of Canberra, Canberra, ACT 2617, Australia

² Research School of Biology, Australian National University, Acton, ACT 2601, Australia

³ Australian Museum Research Institute, Australian Museum, 1 William St Sydney, NSW 2010, Australia

⁴ Institute for Systems Genomics and Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT 06269, USA

⁵ Departament de Biologia Cel·lular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Spain

⁶ Genome Integrity and Instability Group, Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Spain

⁷ Laboratório de Citogenética de Peixes, Departamento de Genética e Evolução, Universidade Federal de São Carlos, São Carlos, SP 13565-905, Brazil

⁸ Graduate School of Pharmaceutical Sciences, Osaka University, Suita 565-0871, Osaka, Japan

⁹ School of Life Sciences, LaTrobe University, Melbourne, VIC 3168, Australia

¹⁰ School of Biosciences, University of Kent, Canterbury CT2 7NJ, UK

¹¹ School of Biological Sciences, The University of Adelaide, Adelaide, SA 5005, Australia

¹² Department of Ecology, Faculty of Science, Charles University, Viničná 7, 128 44 Prague 2, Czech Republic

¹³ Amphibian Research Center, Hiroshima University, Higashi-Hiroshima 739-8526, Japan

¹⁴ Laboratory of Animal Cytogenetics & Comparative Genomics (ACCG), Department of Genetics, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand

¹⁵ School of Natural Sciences, University of Tasmania, Hobart 7000, Australia

* Correspondence: Janine.Deakin@ecanberra.edu.au (J.E.D.); Tariq.Ezaz@canberra.edu.au (T.E.)

† These authors contributed equally to this work.

Received: 18 July 2019; Accepted: 13 August 2019; Published: 20 August 2019



Abstract: The recent advances in DNA sequencing technology are enabling a rapid increase in the number of genomes being sequenced. However, many fundamental questions in genome biology remain unanswered, because sequence data alone is unable to provide insight into how the genome is organised into chromosomes, the position and interaction of those chromosomes in the cell, and how chromosomes and their interactions with each other change in response to environmental stimuli or over time. The intimate relationship between DNA sequence and chromosome structure and function highlights the need to integrate genomic and cytogenetic data to more comprehensively understand the role genome architecture plays in genome plasticity. We propose adoption of the term ‘chromosomics’ as an approach encompassing genome sequencing, cytogenetics and cell biology, and present examples of where chromosomics has already led to novel discoveries, such as the sex-determining gene in eutherian mammals. More importantly, we look to the future and the questions that could be answered as we enter into the chromosomics revolution, such as the role of chromosome rearrangements in speciation and the role more rapidly evolving regions of the genome, like centromeres, play in genome plasticity. However, for chromosomics to reach its full potential, we need to address several challenges, particularly the training of a new generation of cytogeneticists, and the commitment to a closer union among the research areas of genomics, cytogenetics, cell biology and bioinformatics. Overcoming these challenges will lead to ground-breaking discoveries in understanding genome evolution and function.

Keywords: cytogenetics; sex chromosomes; chromosome rearrangements; genome plasticity; centromere; genome biology; evolution

1. Introduction

Advances in technology have made sequencing the entire genome of an organism essentially routine. However, DNA sequence is only one relatively static component of the highly dynamic entities within the nucleus of a cell—chromosomes. Where a particular sequence is located on a chromosome and how it interacts with other parts of the genome are important aspects of genome biology often overlooked in genome sequencing projects. We propose a new framework for studying genome biology that integrates approaches in genome sequencing, cytogenetics and cell biology, as well as a renewed focus on training the next generation of genome biologists in the skills required for the integration of these data. We propose the adoption of the term ‘chromosomics’, which combines the original definition of cytogenetics (chromosomes and cytology) with genomics (gene content, structure and function for an entire organism), to ensure a closer integration of these fields. The term chromosomics was originally proposed by Uwe Claussen to introduce the branch of cytogenetics that deals with the three-dimensional structure of chromosomes and their associated gene regulation [1]. However, we propose that this term encompass the integration of the latest advances in cytogenetics, genome sequencing, epigenomics and cell biology. The adoption of a chromosomics approach to answering the big fundamental questions in biology will undoubtedly lead to major discoveries that were previously beyond reach.

In recent times, the field of genomics has largely distanced itself from cytogenetics, the field providing insight into chromosome structure, function and evolution. This separation of the fields has been to the detriment of a full understanding of how the genome works in the cell. These two fields were never intended to work in isolation. In 1920, Hans Winkler coined the term ‘genome’ to combine the study of genes and chromosomes [2], yet in modern interpretations of ‘genome’, chromosomes are often forgotten and the focus is solely on the DNA sequence. Similarly, Walter Sutton in 1902 (no published record) used the term ‘cytogenetics’ to combine cytology (the study of cell structure and function) with genetics (the study of genes, genetic variations and heredity). However, the cytological aspects of cytogenetics are largely ignored by most modern cytogenetic studies. As these respective fields have narrowed their focus, the result has been the development of technological and methodological advancements (examples in Table 1) that could allow us to more fully capture the dynamic nature and evolution of chromosomes from potentially any species to provide insight into fundamental biological questions.

Table 1. Recent Technological Advances in Cytogenetics and Genomics.

Technique	Uses	Example References
Chromosome microdissection and sequencing	Sequencing individual chromosomes/chromosome segments and haplotyping	[3–5]
Flow sorting of chromosomes and sequencing	Assigning sequences to chromosomes	[6,7]
Fibre-fluorescence in situ hybridisation (FISH)	Ordering of sequences large insert clones on a DNA fibre	[8]
Universal probe set and multiprobe slides	Rapid bacterial artificial chromosome (BAC) mapping across multiple species	[9]

Table 1. Cont.

Homologue-specific oligopaints	Visually distinguish single copy regions of homologous chromosomes	[10]
Super-resolution microscopy	Imaging of chromatin and nuclear organisation	[11]
BioNano	Genome mapping to improve assemblies and detect structural variations	[12]
Long-read sequencing (e.g., PacBio, Oxford Nanopore)	Improving genome assemblies, identifying structural variants	[13–15]
Linked-read sequencing (10X Chromium)	Phasing and improving scaffolding of genome assemblies	[16]
Hi-C sequencing and CHiA-PET	Chromatin interactions and improving genome assemblies (Hi-C)	[17–19]

Chromosomes play a vital role in the nucleus, as they are essential for DNA to replicate and segregate during cell division. They are not randomly positioned in the nucleus, but organised into specific areas called chromosomal territories [20] that change during the cell cycle [21,22] and development [23–25]. Maintenance of these territories is important for proper cell functioning, replication, and the accurate division and differentiation of cells. If we repack the DNA into a chromosome, we see that the DNA is wrapped around a nucleosome consisting of eight histone proteins to produce a chromatin fibre, which is attached to a backbone of non-histone proteins called the chromosome scaffold (Figure 1). The dynamic nature of the chromosome throughout the cell cycle and in response to environmental influences is enabled by the ability of the chromatin fibre to vary the level of DNA compaction and histone composition (epigenetics), and the ability of the scaffold proteins to follow the changes of the chromatin fibre [26]. Chromatin remodelling and changes in chromatin conformation affect interactions between sequences in different genomic regions and can influence gene regulation. The close connection between DNA, chromosome structure and the position of chromosomes in the nucleus highlights the need to integrate genomic data to better understand chromosome architecture and function. This information will provide a more comprehensive understanding of the evolutionary plasticity and organisational functions of genome architecture and mechanisms of faithful transmission of the genome to offspring.

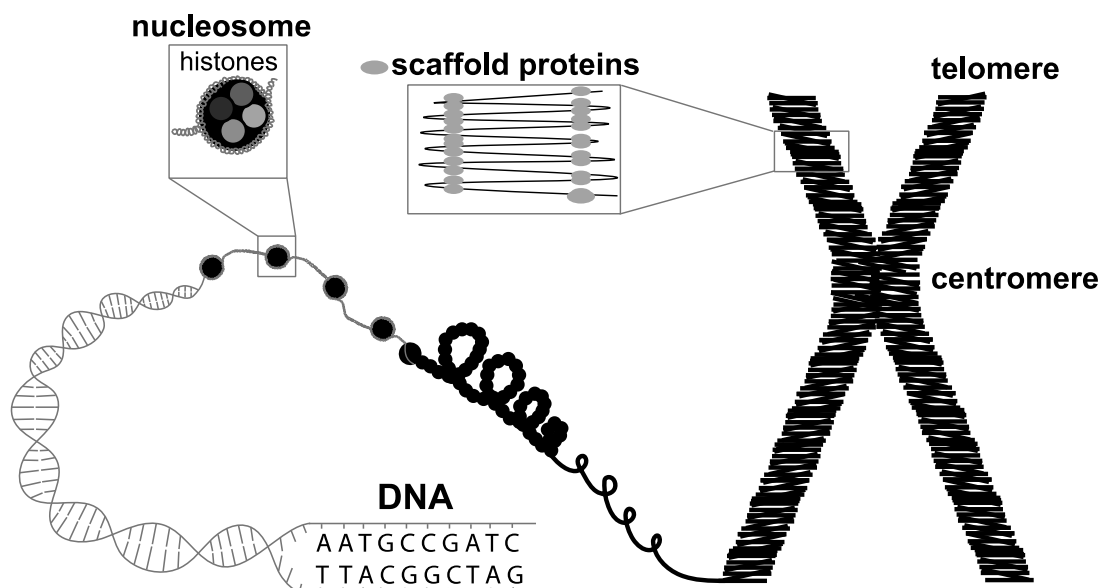


Figure 1. Repacking the DNA into a chromosome. The double-stranded DNA helix is wrapped around a nucleosome consisting of eight histone proteins to produce a chromatin fibre, which is attached to a backbone of non-histone proteins called scaffold proteins which form the chromosome scaffold.

In the past, incremental advances in understanding genome biology have been made through combining information from cytology, cytogenetics and genomics, often from data gathered by different groups focused on one particular question and over many years (e.g., the discovery the Philadelphia chromosome causing chronic myelogenous leukemia or the discovery of the sex-determining gene, *SRY*; Figure 2). Now is the time to reunite cytogenetic and sequencing approaches. Not only has the resolution of chromosomes under various forms of microscopy greatly accelerated (e.g., deconvolution system [27], structured illumination microscopy [26] and super-resolution microscopy [28]), but new sequencing technologies are now promising to make genome assemblies close to chromosome level a reality. In addition, new sequence-based techniques for chromosome conformation capture promise to fill in the details in our cytological picture of how active and inactive chromatin is assembled and arranged into functional units in the interphase nucleus. Collectively, these advances afford the capability to answer key fundamental questions in genome biology.

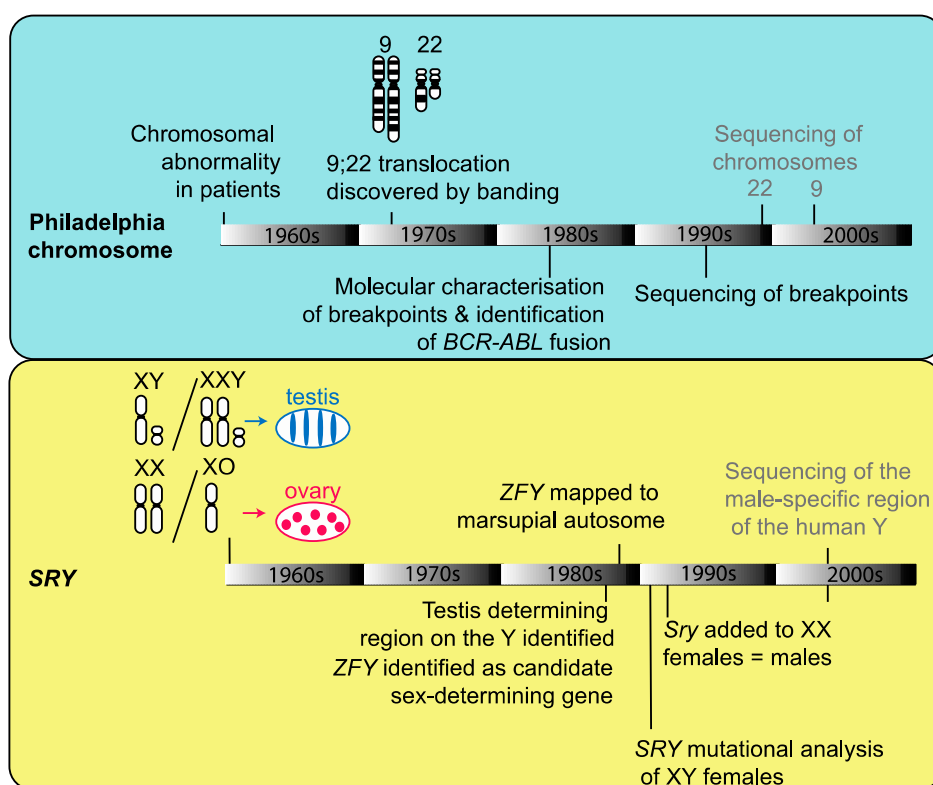


Figure 2. The incremental advances made through combined cytogenetic and genomic information in the discovery of the Philadelphia chromosome causing chronic myelogenous leukemia [29–33] and the discovery of the sex-determining gene *SRY* [34–38].

2. Development of Genome Sequencing from Cytogenetics

The proposal to sequence the human genome led to the start of the genomics era. When we consider the human genome project, it was approached from an understanding that the position of the sequence on the chromosome was important [39]. Indeed, the forerunner of the Human Genome Project was a series of meetings of an international Human Gene Mapping consortium, which met annually to put together increasingly detailed physical maps of all the human chromosomes, combining data from linkage analysis, somatic cell genetics and radiation hybrid analysis, and in situ hybridisation. This consortium was organised into separate committees for each human chromosome, as well as committees for the mitochondrial genome. Moreover, the consortium included a comparative gene mapping committee, which started out largely focused on mouse but grew to encompass many other mammals, birds and fishes. Slowly, the physical maps developed by these working groups expanded

and were filled in with other markers. Sequencing crept in to offer the ultimate detail of individual genes (or at least exomes). The advent of large insert clones like BACs (bacterial artificial chromosomes) greatly aided the extension of DNA sequence to encompass larger genomic intervals beyond individual genes [39].

Physical BAC or yeast artificial chromosome (YAC) maps of each chromosome were constructed and sequenced, resulting in a chromosome-based genome assembly and enabling the integration of gene and genetic mapping data accumulated over many years and by many different researchers [40]. The subsequent ENCODE (Encyclopedia of DNA elements) project saw the integration of sequence data with information on chromatin states, which provided an exceptional insight into dynamic gene regulation [41]. However, just as genome sequences for other species were needed to help interpret the human genome [42], comparative data from a broad range of species are required to fully understand the role many chromatin modifications play in genome function. The interpretation of the ENCODE data for the human genome was only made possible by the chromosome-based genome assembly, affording an appreciation for regional transcriptional control, dynamic chromatin states and long-range interlocus interactions. At present, a challenge for genomes from non-traditional model species is the difficulty in overlaying chromatin remodelling data when genome assemblies are not yet at the chromosome level. The platypus (*Ornithorhynchus anatinus*) genome is an excellent example of the difficulty in accurately overlaying and interpreting DNA methylation data on a fragmented genome assembly. The platypus genome was sequenced to approximately six-fold coverage by a whole genome shotgun approach using Sanger sequencing, and only around 21% of this genome assembly was anchored to platypus chromosomes [43]. Although a valuable resource, the low percentage of the genome anchored to chromosomes greatly reduced the number of genes that could be examined in a recent comparative study of reduced representation bisulphite sequencing data [44].

3. The Integral Role of Cytogenetics in Genome Projects

With increasingly cost-effective high throughput sequencing, most recently assembled genomes feature short contigs and often lack even a basic physical map or chromosome number and morphology information. While chromosome-level assemblies might not be feasible for all genomes targeted for sequencing, they should be well represented across all lineages to allow comparative genome biology studies that, by their very nature, rely on knowing the position of orthologous sequences among genomes. All genome sequencing projects should incorporate some level of cytogenetic analysis from the very start. For example, a logical first step in any whole genome sequence project would be to ensure that the individual being sequenced is not carrying chromosomal aberrations, particularly if the genome assembly is to be used as a reference for population-level sequencing. Basic karyotyping would ensure the ploidy level of the species, the absence of aneuploidy and confirm the genetic sex of the individual when cytogenetically distinguishable sex chromosomes are present, a particularly important consideration in species subject to environmental sex reversal (e.g., *Pogona vitticeps* [45]). Karyotyping will also determine if there are large heterochromatic chromosomes or regions. The flow sorting of chromosomes can be used for gross assessment of aneuploidy. Flow cytometry with appropriate standards is a reliable and fast method for estimating genome size [46], an important consideration in determining the amount of sequencing required to achieve the desired level of genome assembly.

A whole genome sequence is much more informative if it is assigned and oriented onto chromosomes, and is far more intuitive to visualise as chromosomes than unconnected and unordered scaffolds. When whole genome sequences fall short of this 'chromosome level assembly', their use for critical aspects of evolutionary and applied biology is significantly limited. Assigning sequence contigs to chromosomes has most often been achieved by integrating sequence data with molecular cytogenetic mapping data. This can be achieved by determining the location of a large-insert clone by fluorescence in situ hybridisation (FISH) on metaphase chromosomes or even extended chromatin fibres (fibre FISH), facilitating physical fine mapping of contigs. In Sanger sequenced genomes, this was accomplished by assigning BAC clones corresponding to individual, large sequence scaffolds. For example, the opossum

genome, with a scaffold N50 (a measurement of assembly quality where 50% of scaffolds are this size or larger) of 59.8 Mb and 97% of the sequence contained in 216 scaffolds, was anchored onto the eight opossum autosomes and the X chromosome by mapping 415 BAC clones [47,48]. However, the proportion of the genome assembly assigned to chromosomes is dependent on the quality of the genome (i.e., N50 size and number of scaffolds). The platypus genome is a prime example, where the high repeat content resulted in a scaffold N50 of 957 kb, and thus only about 21% of the genome was chromosome-anchored [43]. An excellent example of where a cytogenetics approach vastly improved the accuracy of genome assembly is the tomato genome. The tomato genome, sequenced by a combination of Sanger and next generation sequencing technologies [49], benefitted greatly from the physical assignment of sequence scaffolds of BACs by FISH and confirmation by optical mapping [50]. The original tomato assembly was ordered based on a high-density linkage map. Differences in arrangement between the linkage and cytogenetic/optical maps were detected for one-third of these scaffolds, mainly in pericentric regions where a reduced level of recombination renders linkage mapping less reliable [50]. The benefits gained from assigning even a portion of the sequence to chromosomes are immense, as highlighted by the chromosomics successes listed in Table S1.

Many genomes sequenced over the past decade have used a ‘shotgun’ approach based on short read sequence technologies to produce a series of scaffolds, often several hundred per chromosome, which are neither anchored to, nor ordered on, the chromosomes. Anchoring every scaffold to a chromosome would be a labour-intensive task, particularly if the assembly has a higher number of scaffolds; by combining computational approaches to merge scaffolds with either cytogenetic mapping and/or PCR-based scaffold verification, chromosome-level assemblies are a more achievable exercise [9,51]. In addition, the development of universal BAC clone probe sets that can be used in a high-throughput, cross-species, multiple hybridisation approach are speeding up the process of developing cytogenetic maps [9]. The advances in sequencing technology that are producing more contiguous genome assemblies, such as the contact sequencing approach of HiC-seq (e.g., Dovetail) [17], linked-read sequencing approach (10X Genomics) [12,16], long read technologies like PacBio [13] and Oxford Nanopore [15], and optical mapping (BioNano) [12], combined with high-throughput cytogenetic methods, will place chromosome-level assemblies within reach for many species.

4. The Big Questions in Genome Biology Requiring a Chromosomics Approach

Despite the meticulous approach taken for the human genome, gaps remained in the genome sequence when the ‘finished’ euchromatic sequence of the human genome was published in 2004 [52]. These ‘black holes’ of the genome corresponded to the most repetitive regions such as, but not limited to, the critically important centromeres [53], nucleolar organiser regions (NORs) [54] and the Y chromosome [36]. Repetitive regions are some of the most rapidly evolving sequences and therefore, are among the most interesting regions of the genome. By employing a chromosomics approach, these hot spots of evolution are beginning to lose their black hole status in the human genome, as well as in other species. We discuss the fundamental questions arising from these evolutionary dynamic regions in relation to two overarching themes associated with genome evolution: genome plasticity and sex chromosome evolution.

4.1. Genome Plasticity and Chromosome Evolution

Why do species have specific karyotypes? Why do chromosome numbers vary greatly within some groups, but are largely the same in others? Why are some of the regions of the genome so well conserved? Why are genomes so extensively changed among closely related species and others strongly conserved? Why do some chromosome rearrangements appear to lead to speciation, yet others are tolerated within a species or population of species? Is the underlying mechanism responsible for chromosomal speciation the same as that leading to chromosomal rearrangements in a disease context (i.e., cancer)? These are fundamental questions regarding genome plasticity that remain unanswered,

and a chromosomics approach is essential for major breakthroughs. The answers to these questions will have wide-ranging impacts in the field of biology. Unlocking the genomic basis of speciation is a biological research priority, fuelled by the ongoing debate on species concepts and facilitated by the availability of an unprecedentedly large number of genomic resources.

The concept of chromosomal speciation, at one stage considered a major contributor in separating populations that differ by a structural rearrangement, was virtually abandoned in favour of theories of a gradual accumulation of mutations in 'speciation genes' (e.g., Reference [55]). The implementation of the most recent 'suppressed recombination model' [56,57] has now fuelled the field using a combination of sequence and cytogenetics [58–60]. In this context, chromosome rearrangements could have a minimal influence on fitness, but would suppress recombination, leading to the reduction of gene flow across genomic regions and to the accumulation of incompatibilities.

Understanding chromosomal speciation is also critical to determining the mechanism(s) underlying genome adaptation to environmental factors and how biodiversity is generated and transmitted to subsequent generations. With so many threatened species across the globe, understanding why some structural variants are tolerated within a population while others lead to reproductive isolation could prove important for the management of breeding programs for species conservation programs. In a disease context, a greater knowledge of the drivers of genome instability will aid research into human and animal diseases, particularly cancers.

Regions of genome instability can have dramatic effects for an organism. Despite being the subject of many studies using a range of species, the underlying molecular mechanisms resulting in genome restructuring/reshuffling are relatively poorly understood. For example, it remains unclear if chromosomal changes associated with speciation arise because there is an adaptive value to a specific chromosomal configuration, and what causes the genomic instability in the first place. The combined use of comparative genomics and cytogenetics of both closely and distantly related mammalian species has been extremely useful in defining models that explain genome structure and evolution [61–67]. Such reconstructions have revealed that the genomic regions implicated in structural evolutionary changes disrupt genomic synteny (evolutionary breakpoint regions, EBRs) and are clustered in regions more prone to breaking and reorganisation [61–63]. In searching for the origin (and consequences) of this evolutionary instability, approaches based purely on genome sequence have only revealed that EBRs are enriched for repetitive sequences, including segmental duplications and transposable elements, which provide the templates for non-allelic homologous recombination, resulting in inversions and additional structural changes [68,69]. Likewise, repetitive sequences in centromeric regions have been implicated in illegitimate recombination events forming Robertsonian fusions [62,70,71]. EBRs also typically occur in gene-dense regions, enriched with genes involved in adaptive processes, where changes to gene expression caused by a chromosomal rearrangement may provide a selective advantage [60,63,72]. Consequently, given the diversity of factors associated with EBRs, it is unlikely that the sequence composition of genomes is solely responsible for genomic instability during evolution and speciation.

As chromosomes are more than just DNA sequence, a more comprehensive approach that incorporates global genomic information on recombination rates, chromatin accessibility, gene function data and nuclear architecture is providing more insight into the factors underpinning genome instability. Of course, chromosome-level assemblies are an essential resource for accurate interpretation of the combination of all these data because, without such assemblies, we have a very limited (if any) understanding of the extent of the genomic restructuring that may have occurred between species, or between normal and disease states, that facilitated changes in global genomic features.

Furthermore, it has become clear that the interplay between the organisation of the genome and nuclear architecture is central to genome function [73]. We have seen a rapid evolution of methods by which to analyse genome organisation and nuclear architecture, moving from cytogenetic approaches, providing a direct measurement within individual cells of distances between loci, to chromosome conformation capture approaches (3C, 4C, 5C and Hi-C), which infer the contact among loci, typically in populations of cells rather than single cells [74,75]. However, comparison of the results obtained

from chromosome conformation capture methods and FISH analyses demonstrates that care needs to be taken when interpreting data obtained solely by one method, suggesting that the use of a combined molecular and cytogenetic approach will lead to more accurate 3D models of genome organisation [76,77].

A new model for genome rearrangements, referred to as the integrative breakage model, has recently been proposed, bringing together all of these features [64]. It posits that genome reshuffling permissiveness is influenced by (i) the physical interaction of genomic regions inside the nucleus, (ii) the accessibility of chromatin states and (iii) the maintenance of essential genes and/or their association with long-range cis-regulatory elements (Figure 3). An initial test of the integrative breakage model using rodents has supported this model [65]. EBRs were found to not only coincide with regions enriched for repetitive sequences and genes, especially genes involved in reproduction and pheromone detection, but possessed the characteristics of open, actively transcribed chromatin. The challenge remains to use a broader spectrum of species to fully test this model and dissect the underlying mechanism for chromosomal rearrangements. Such studies will now be possible with the ability to achieve chromosome-level assemblies and obtain information on chromatin modifications and nuclear architecture for non-traditional model species. A similar approach could be extended to intraspecific comparisons, such as normal versus disease samples or samples across a population where structural variants are known.

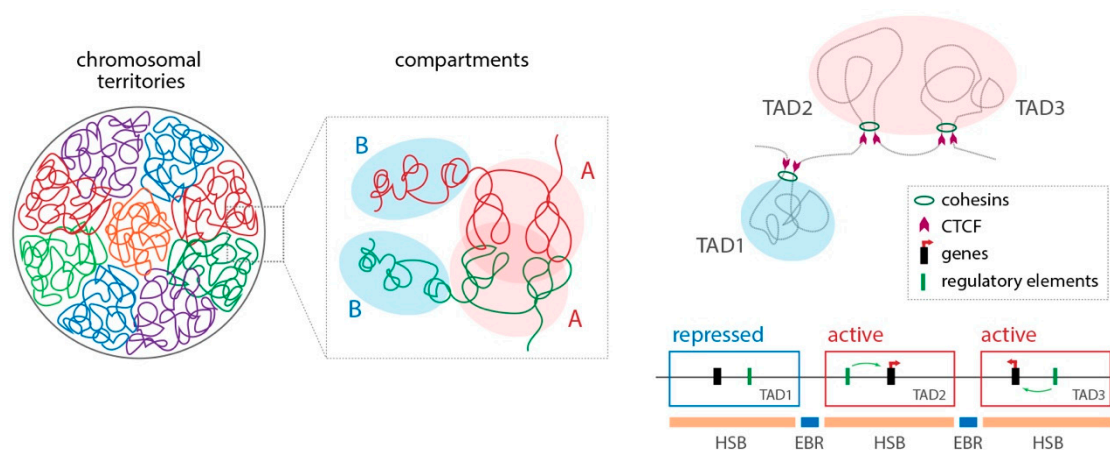


Figure 3. The integrative breakage model, a multilayer framework for the study of genome evolution that takes into account the high-level structural organisation of genomes and the functional constraints that accompany genome reshuffling [64]. Genomes are compartmentalised into different levels of organisation that include: (i) chromosomal territories, (ii) ‘open’ (termed ‘A’)/‘closed’ (termed ‘B’) compartments inside chromosomal territories, (iii) topologically associated domains (TADs) and (iv) looping interactions. TADs, which are delimited by insulating factors such as CTCF and cohesins, harbour looping topologies that permit long-range interactions between target genes and their distal enhancers, thus providing ‘regulatory neighbourhoods’ within homologous syntenic blocks (HSBs). In this context, the integrative breakage model proposes that genomic regions involved in evolutionary reshuffling (evolutionary breakpoint regions, EBRs) which will likely be fixed within populations are (i) those that contain open chromatin DNA configurations and epigenetic features that could promote DNA accessibility and therefore genomic instability, and (ii) that do not disturb essential genes and/or gene expression.

4.2. Sex Chromosome Evolution: Genetics and Epigenetics

Sex chromosomes represent one of the most dynamic parts of any genome, as they are highly variable in morphology and sequence content across the plant and animal kingdoms. The special evolutionary forces experienced by sex chromosomes have rendered them highly complex entities within the genome; thus, it remains a challenge for evolutionary biologists to disentangle the varied mechanisms involved in their evolution. There are still many fundamental questions that remain

unanswered because of the genomic black hole status of sex chromosomes. Why do sex chromosomes evolve and degenerate in some species but not in others? Why do sex chromosomes have a propensity to accumulate repetitive sequences that, in most cases, are species-specific? How do sex chromosomes drive speciation and hybrid incompatibilities? Why do sex chromosomes vary within a species? Why do some species have complete dosage compensation mechanisms while others do not? The complexity of sex chromosome origin, evolution and gene organisation is multilayered and cannot be understood by studying a single aspect of its biology alone. Therefore, a chromosomics approach, taking into account all aspects of cellular and molecular biology, will be essential to answering these questions.

Many genome projects were undertaken without considering the sex chromosomes, with most projects intentionally choosing to sequence the homogametic sex in order to obtain higher sequence coverage and better assembly of the X or Z chromosome, completely neglecting to obtain sequence for the Y or W, thus ignoring the complexities of sex-delimited sex chromosome variation. Simple karyotyping with basic banding analysis, or painting one sex chromosome onto the other, can be very informative about the DNA content of the sex chromosomes. Such experiments can provide valuable information to support the adoption of appropriate sequencing technologies to obtain sequences from those unique but difficult to sequence regions of the genome. In a recent review, Tomaszewicz et al. [78] highlighted the need to sequence sex chromosomes, and elegantly described challenges and opportunities for combining new and emerging technologies to sequence these difficult regions of the genome. Only a chromosomics approach, combining cytogenetics and appropriate sequencing platform(s), can answer the fundamental questions regarding sex chromosome evolution. As an example, a human Y chromosome of African origin was recently assembled by flow sorting nine million Y chromosomes and sequencing using the Oxford Nanopore MinION platform, resulting in a Y chromosome assembly with an N50 of 1.46Mb [79], yet this method was much more time- and cost-efficient than that used to obtain the original human Y chromosome sequence [36,80].

Determining the epigenetic status of the sex chromosomes and the genes they contain is also extremely valuable. For example, the Chinese half-smooth tongue sole (*Cynoglossus semilaevis*) is a species with genetic sex determination (ZZ males and ZW females), but with a temperature override mechanism, where exposure of developing embryos to high temperatures causes genetic ZW females to develop as males (sex reversal). The sex determining gene *dmrt1* [81] is epigenetically silenced by DNA methylation in ZW females, but not in sex-reversed ZW males, where *dmrt1* expression is upregulated, leading to initiation of the male development pathway [82].

Dosage compensation, a mechanism equalizing the expression of genes on the sex chromosomes between males and females, is epigenetically controlled. A comparison of the gene content of the X chromosomes of eutherians and marsupials would suggest that a dosage compensation mechanism, in the form of X chromosome inactivation, may be shared between these two mammalian groups, as the X chromosome of marsupials is homologous to approximately two-thirds of the X of eutherians [83]. However, epigenetic analyses point to an independent evolutionary origin of X chromosome inactivation in marsupials and eutherians [84]. Similarly, there are striking differences in the extent and mechanisms of dosage compensation between more divergent taxa. For example, *Drosophila melanogaster* increases X chromosome transcription by the binding of Male Specific Lethal (MSL) complex to the single X chromosome in males to achieve dosage compensation [85]. In contrast, many species, including insects, fishes, birds, reptiles and platypus, have incomplete dosage compensation [86]. Reports of incomplete dosage compensation have most often relied purely on a sequence-based approach to measure the average transcriptional output of the X or Z chromosome between males and females for a population of cells, which does not afford an understanding of the mechanisms that impact differential transcription. However, examination of individual cells and measures of gene copies from the two Z or X chromosomes using a technique detecting nascent transcription (RNA-FISH) provides information on a single cell basis. For example, in the homogametic sex of chicken (*Gallus gallus*) and platypus, RNA-FISH detected a portion of cells expressing a gene from one copy of the X/Z, while a portion was

expressed from both copies, which explains the incomplete dosage compensation pattern observed by transcriptome approaches measuring population of cells [87–89].

Sex chromosomes also have an impact beyond simply facilitating sex determination. In *Drosophila*, for example, polymorphisms in repetitive sequences on the Y chromosome influence gene expression of genes across the genome, particularly those involved in chromosome organisation and chromatin assembly [90,91]. Essentially, the polymorphic Y chromosome is a source of epigenetic variation in *Drosophila*. This epigenetic variation has implications for speciation. Engineered species hybrids showed either reduced fertility or rescued fertility depending on the origin of the Y chromosome and grandparental genetic background of the hybrid, suggesting that the Y chromosome may contribute to reproductive isolation [92]. The regulatory effect of the Y chromosome on gene expression is not limited to *Drosophila*, but has been demonstrated, at least for immune-related genes, in humans and mice [93,94]. This regulatory role for the Y highlights the importance of ascertaining not only the DNA sequence, but also the epigenetic status of sex chromosomes, in addition to, and in the context of, the rest of the genome.

5. Challenges Ahead for Chromosomics

Chromosomics approaches can have far greater success for answering fundamental biological questions than either genomics or cytogenetics approaches alone; this begs the question: why haven't these two fields merged more extensively? We have identified three major challenges that may be preventing a closer union of these fields.

The biggest challenge for chromosomics is the dwindling number of researchers worldwide with expertise in cytogenetics. Rejuvenating the training of cytogeneticists is essential if the potential of chromosomics as a field is to reach its full potential. At a "Cytogenetics in the Genomics Era" workshop held in 2017 at the University of Canberra, we identified a need to renew excitement in chromosomes among undergraduate and graduate students worldwide. Genetics and genomics courses are often taught by those who have little experience in or appreciation for chromosomes, perhaps leading to anxiety in students around the study of chromosome biology. The origin of this mismatch can also be found in the backgrounds of leaders in the fields of DNA sequencing and of cytogenetics. Genome sequencing researchers often have backgrounds in biochemistry, and their training may be entirely devoid of exposure to genetics. In contrast, those who gravitate to chromosome work often have a background of zoology or botany, and may perhaps have had little exposure to biochemistry. At a time when we are more aware than ever before of the important role genome organisation and nuclear architecture plays in genome function, it is imperative that students gain an understanding of, and appreciation for, the basics of chromosome biology from their first introduction to the world of genetics and genomics. This early introduction needs to be followed by reinforcement throughout their studies. Some good courses in integrated cell and molecular biology would be a step in the right direction.

Furthermore, graduate students have been attracted to the rapidly advancing world of genome sequencing, where mountains of data are now being rapidly and cheaply generated, as opposed to cytogenetics projects, where data are accumulated more slowly. We need to instil in students the incredible experience of observing the amazing structure of chromosomes under a microscope, and the importance of understanding chromosome structure and continuing to develop more high-throughput approaches to cytogenetics to keep pace with the rapidly advancing world of genome technology. The genomics field over the past decade has made major technological advances in obtaining genome sequences faster and more cheaply, driven mostly by the large community in this field. Advancements of a similar level in cytogenetics will require a larger community of researchers to drive the need for the technology. By increasing the training of researchers in cytogenetics, we not only increase the uptake of chromosomics, but generate a potential pool of people able to develop technologies to achieve cytogenetic analyses fast and cheaper.

A challenge for cytogeneticists is that chromosome work, which used to be the cheapest aspect of a genome project, is now the most expensive. New sequencing technology has brought down the cost of sequencing by six orders of magnitude, and the speed at which data are generated has dramatically increased. Technical innovation in cell biology, in contrast, has greatly magnified costs. High throughput really does not apply to chromosome observation or experimentation.

The second most challenging skills area for chromosomics is the need for sophisticated bioinformatics. Whole genomes can now be rapidly sequenced, but assembling the sequence is much slower, and more labour- and computationally-intensive. Likewise, overlaying genome sequence with 3D chromatin structure data often presents a computational challenge [95]. More importantly, the incorporation of cytogenetic information with genomic data is not commonly attempted. For chromosomics approaches to be more readily applied in the future, we require more bioinformaticians to be trained generally in genome assembly, as well as with an appreciation for cytogenetics.

Another challenge is ensuring samples, whether from wild species, laboratory species or clinical specimens, are collected appropriately for cytogenetic analysis. The collection of samples for DNA analysis is now routine, but the collection of material for the combination of cytogenetic and genomic analysis is not. The special requirements of samples collected for cytogenetic analysis need to be disseminated more widely. This is a relatively easy challenge to address by making a 'field guide to chromosomics' available to field researchers and a similar one to those working in a clinical setting, detailing how samples should be collected and stored for the implementation of chromosomic techniques. With the appropriate samples, a chromosomics approach could be employed to study structural and epigenetic variation at population or biogeography levels, where there is the potential to uncover the underlying genetic or epigenetic basis for adaptation to a particular environment. These data could have an impact on the conservation and management of threatened species and lead to a greater understanding of the factors underlying disease phenotypes.

6. Future Opportunities

6.1. Sequencing Genome Black Holes

Resolving highly repetitive regions of genomes, i.e., the black holes, is now possible, and we will soon have the capability to explore these previously under-represented regions of the genome to more fully understand their evolution and function. An important step towards understanding the evolution and function of repetitive regions has been the development of tools able to analyse and visualise these regions of the genome, such as RepeatExplorer [96]. Furthermore, sequencing of highly repetitive regions is now possible with long-read sequencing technology. For example, the centromeric sequence of the human Y chromosome has recently been assembled [97], demonstrating that nanopore long technology has the potential to fill in genome black holes. Likewise, combinations of different approaches where individual sex chromosomes are sequenced are proving successful in resolving at least the non-repetitive regions of sex chromosomes and identifying candidate sex-determining genes [98].

6.2. Spatial Chromosome Organisation

The importance of the territorial organisation of chromosomes (chromosome territories) in plant and animal cells was proposed over a century ago by several cytologists (reviewed in Reference [3]). We are currently at a stage where the capacity to study the changes in spatial organisation in a population of cells, or even a single cell, is possible. The recent advances in chromatin analysis, coupled with next generation sequencing (e.g., Hi-C, Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET)) and 3D and 4D FISH, live-cell and super-resolution microscopy, provide opportunities to garner a more comprehensive understanding of chromosomal activities within the nucleus contributing to gene regulation, expression and ultimate phenotypic outcomes of an individual [22–25,99]. Such a

chromosomics approach is already underway for the human genome with the launch of the 4D Nucleome project to understand how the changes in chromosome dynamics contribute to gene regulation in different cell types and biological states [99]. We will gain unprecedented insight into the role of the spatial organisation of chromosomes in genome evolution by extending this same approach to many more species.

7. Concluding Remarks

We are on the verge of an exciting new revolution in biology, with a change from thinking of genomes as one-dimensional entities to defining the ways every component of the genome is packaged and changes through space and time. Chromosomics is the best path forward, providing one comprehensive analysis to answer complex questions in evolution and disease contexts. However, this can only be achieved if genomicists, cytogeneticists, cell biologists and bioinformaticians commit to forming a closer union for advancing this new era in genome biology.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4425/10/8/627/s1>. Table S1: Examples of research questions answered with a chromosomics approach.

Author Contributions: The concept and idea of this review was developed during the workshop “Cytogenetics in the Genomics Era” organised by T.E. and J.E.D. at the Institute for Applied Ecology, University of Canberra, February 2017. J.E.D. and T.E. prepared the first draft after S.P., R.O., A.R.-H., M.B.C., M.D.B.E., K.F., J.A.M.G., D.G., F.G., L.K., I.M., M.R., K.S. and E.W. contributed to drafting the manuscript outline. JED, S.P., R.O, A.R.-H., M.B.C., M.D.B.E., K.F., J.A.M.G., D.G., F.G., L.K., I.M., M.R., K.S., E.W. and T.E. contributed to writing and revising drafts of the manuscripts. J.E.D., S.P. and A.R.-H. prepared figures. All authors approved the final version.

Funding: The workshop was funded by Institute for Applied Ecology, University of Canberra strategic funds awarded to T.E. and J.E.D.

Acknowledgments: We thank Arthur Georges, Craig Moritz, Stephen Sarre who contributed to discussions as part of the “Cytogenetics in the Genomics Era” workshop.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Claussen, U. Chromosomics. *Cytogenet. Genome Res.* **2005**, *111*, 101–106. [[CrossRef](#)] [[PubMed](#)]
2. Winkler, H. *Verbreitung und Ursache der Parthenogenesis im Pflanzen-und Tierreiche*; Verlag von Gustav Fischer: Jena, Germany, 1920.
3. Traut, W.; Vogel, H.; Glöckner, G.; Hartmann, E.; Heckel, D.G. High-throughput sequencing of a single chromosome: A moth W chromosome. *Chromosom. Res.* **2013**, *21*, 491–505. [[CrossRef](#)] [[PubMed](#)]
4. Seifertova, E.; Zimmerman, L.B.; Gilchrist, M.J.; Macha, J.; Kubickova, S.; Cernohorska, H.; Zarsky, V.; Owens, N.D.L.; Sesay, A.K.; Tlapakova, T.; et al. Efficient high-throughput sequencing of a laser microdissected chromosome arm. *BMC Genom.* **2013**, *14*, 1. [[CrossRef](#)] [[PubMed](#)]
5. Ma, L.; Li, W.; Song, Q. Chromosome-range whole-genome high-throughput experimental haplotyping by single-chromosome microdissection. *Methods Mol. Biol.* **2017**, *1551*, 161–169.
6. Makunin, A.I.; Kichigin, I.G.; Larkin, D.M.; O'Brien, P.C.M.; Ferguson-Smith, M.A.; Yang, F.; Proskuryakova, A.A.; Vorobieva, N.V.; Chernyaeva, E.N.; O'Brien, S.J.; et al. Contrasting origin of B chromosomes in two cervids (Siberian roe deer and grey brocket deer) unravelled by chromosome-specific DNA sequencing. *BMC Genom.* **2016**, *17*, 618. [[CrossRef](#)] [[PubMed](#)]
7. Murchison, E.P.; Schulz-Trieglaff, O.B.; Ning, Z.; Alexandrov, L.B.; Bauer, M.J.; Fu, B.; Hims, M.; Ding, Z.; Ivakhno, S.; Stewart, C.; et al. Genome sequencing and analysis of the Tasmanian devil and its transmissible cancer. *Cell* **2012**, *148*, 780–791. [[CrossRef](#)] [[PubMed](#)]
8. Skinner, B.M.; Sargent, C.A.; Churcher, C.; Hunt, T.; Herrero, J.; Loveland, J.E.; Dunn, M.; Louzada, S.; Fu, B.; Chow, W.; et al. The pig X and Y Chromosomes: Structure, sequence, and evolution. *Genome Res.* **2016**, *26*, 130–139. [[CrossRef](#)] [[PubMed](#)]
9. Damas, J.; O'Connor, R.; Farré, M.; Lenis, V.P.E.; Martell, H.J.; Mandawala, A.; Fowler, K.; Joseph, S.; Swain, M.T.; Griffin, D.K.; et al. Upgrading short read animal genome assemblies to chromosome level using comparative genomics and a universal probe set. *Genome Res.* **2017**, *27*, 875–884. [[CrossRef](#)] [[PubMed](#)]

10. Beliveau, B.J.; Boettiger, A.N.; Avendaño, M.S.; Jungmann, R.; McCole, R.B.; Joyce, E.F.; Kim-Kiselak, C.; Bantignies, F.; Fonseka, C.Y.; Erceg, J.; et al. Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nat. Commun.* **2015**, *6*, 7147. [[CrossRef](#)]
11. Lakadamyali, M.; Cosma, M.P. Advanced microscopy methods for visualizing chromatin structure. *FEBS Lett.* **2015**, *589*, 3023–3030. [[CrossRef](#)]
12. Mostovoy, Y.; Levy-Sakin, M.; Lam, J.; Lam, E.T.; Hastie, A.R.; Marks, P.; Lee, J.; Chu, C.; Lin, C.; Džakula, Ž.; et al. A hybrid approach for de novo human genome sequence assembly and phasing. *Nat. Methods* **2016**, *13*, 12–17. [[CrossRef](#)] [[PubMed](#)]
13. Rhoads, A.; Au, K.F. PacBio Sequencing and Its Applications. *Genom. Proteom. Bioinform.* **2015**, *13*, 278–289. [[CrossRef](#)] [[PubMed](#)]
14. Magi, A.; Semeraro, R.; Mingrino, A.; Giusti, B.; D'Aurizio, R. Nanopore sequencing data analysis: State of the art, applications and challenges. *Brief. Bioinform.* **2017**, *19*, 1256–1272. [[CrossRef](#)] [[PubMed](#)]
15. Lu, H.; Giordano, F.; Ning, Z. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genom. Proteom. Bioinform.* **2016**, *14*, 265–279. [[CrossRef](#)] [[PubMed](#)]
16. Zheng, G.X.Y.; Lau, B.T.; Schnall-Levin, M.; Jarosz, M.; Bell, J.M.; Hindson, C.M.; Kyriazopoulou-Panagiotopoulou, S.; Masquelier, D.A.; Merrill, L.; Terry, J.M.; et al. Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nat. Biotechnol.* **2016**, *34*, 303–311. [[CrossRef](#)] [[PubMed](#)]
17. Putnam, N.H.; Connell, B.O.; Stites, J.C.; Rice, B.J.; Hartley, P.D.; Sugnet, C.W.; Haussler, D.; Rokhsar, D.S. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* **2016**, *26*, 342–350. [[CrossRef](#)] [[PubMed](#)]
18. Dudchenko, O.; Batra, S.S.; Omer, A.D.; Nyquist, S.K.; Hoeger, M.; Durand, N.C.; Shamim, M.S.; Machol, I.; Lander, E.S.; Aiden, A.P. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **2017**, *10*, 92–95. [[CrossRef](#)] [[PubMed](#)]
19. Fullwood, M.J.; Liu, M.H.; Pan, Y.F.; Liu, J.; Xu, H.; Mohamed, Y.B.; Orlov, Y.L.; Velkov, S.; Ho, A.; Mei, P.H.; et al. An oestrogen-receptor- α -bound human chromatin interactome. *Nature* **2009**, *462*, 58–64. [[CrossRef](#)] [[PubMed](#)]
20. Cremer, T.; Cremer, C. Rise, fall and resurrection of chromosome territories: A historical perspective. Part II. Fall and resurrection of chromosome territories during the 1950s to 1980s. Part III. Chromosome territories and the functional nuclear architecture: Experiments and models from the 1990s to the present. *Eur. J. Histochem.* **2006**, *50*, 223–272.
21. Gibcus, J.H.; Samejima, K.; Goloborodko, A.; Samejima, I.; Naumova, N.; Nuebler, J.; Kanemaki, M.T.; Xie, L.; Paulson, J.R.; Earnshaw, W.C.; et al. A pathway for mitotic chromosome formation. *Science* **2018**, *359*, eaao6135. [[CrossRef](#)]
22. Naumova, N.; Imakaev, M.; Fudenberg, G.; Zhan, Y.; Lajoie, B.R.; Mirny, L.A.; Dekker, J. Organization of the mitotic chromosome. *Science* **2013**, *342*, 948–953. [[CrossRef](#)]
23. Patel, L.; Kang, R.; Rosenberg, S.C.; Qiu, Y.; Raviram, R.; Chee, S.; Hu, R.; Ren, B.; Cole, F.; Corbett, K.D. Dynamic reorganization of the genome shapes the recombination landscape in meiotic prophase. *Nat. Struct. Mol. Biol.* **2019**, *26*, 164–174. [[CrossRef](#)] [[PubMed](#)]
24. Alavattam, K.G.; Maezawa, S.; Sakashita, A.; Khoury, H.; Barski, A.; Kaplan, N.; Namekawa, S.H. Attenuated chromatin compartmentalization in meiosis and its maturation in sperm development. *Nat. Struct. Mol. Biol.* **2019**, *26*, 175–184. [[CrossRef](#)] [[PubMed](#)]
25. Vara, C.; Paytuví-Gallart, A.; Cuartero, Y.; Le Dily, F.; Garcia, F.; Salvà-Castro, J.; Gómez-H, L.; Julià, E.; Moutinho, C.; Aiese Cigliano, R.; et al. Three-dimensional genomic structure and cohesin occupancy correlate with transcriptional activity during spermatogenesis. *Cell Rep.* **2019**, *28*, 352–367. [[CrossRef](#)] [[PubMed](#)]
26. Poonperm, R.; Takata, H.; Hamano, T.; Matsuda, A.; Uchiyama, S.; Hiraoka, Y.; Fukui, K. Chromosome scaffold is a double-stranded assembly of scaffold proteins. *Sci. Rep.* **2015**, *5*, 11916. [[CrossRef](#)] [[PubMed](#)]
27. Wako, T.; Fukuda, M.; Furushima-Shimogawara, R.; Belyaev, N.D.; Turner, B.M.; Fukui, K. Comparative analysis of topographic distribution of acetylated histone H4 by using confocal microscopy and a deconvolution system. *Anal. Chim. Acta* **1998**, *365*, 9–17. [[CrossRef](#)]
28. Nir, G.; Farabella, I.; Pérez Estrada, C.; Ebeling, C.G.; Beliveau, B.J.; Sasaki, H.M.; Lee, S.H.; Nguyen, S.C.; McCole, R.B.; Chatteraj, S.; et al. Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLoS Genet.* **2018**, *14*, 1–35. [[CrossRef](#)] [[PubMed](#)]

29. Nowell, P.C.; Hungerford, D.A. Chromosome studies on normal and leukemic human leukocytes. *J. Natl. Cancer Inst.* **1960**, *25*, 85–109.
30. Rowley, J.D. A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **1973**, *243*, 290. [[CrossRef](#)]
31. Shtivelman, E.; Lifshitz, B.; Gale, R.P.; Canaani, E. Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. *Nature* **1985**, *315*, 550–554. [[CrossRef](#)]
32. Humphray, S.J.; Oliver, K.; Hunt, A.R.; Plumb, R.W.; Loveland, J.E.; Howe, K.L.; Andrews, T.D.; Searle, S.; Hunt, S.E.; Scott, C.E.; et al. DNA sequence and analysis of human chromosome 9. *Nature* **2004**, *429*, 369–374. [[CrossRef](#)] [[PubMed](#)]
33. Dunham, I.; Shimizu, N.; Roe, B.A.; Chissoe, S.; Dunham, I.; Hunt, A.R.; Collins, J.E.; Bruskiwich, R.; Beare, D.M.; Clamp, M.; et al. The DNA sequence of human chromosome 22. *Nature* **1999**, *402*, 489–495. [[CrossRef](#)] [[PubMed](#)]
34. Foster, J.W.; Graves, J.A.M. An SRY-related sequence on the marsupial X chromosome: Implications for the evolution of the mammalian testis-determining gene. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 1927–1931. [[CrossRef](#)] [[PubMed](#)]
35. Sinclair, A.H.; Berta, P.; Palmer, M.S.; Hawkins, J.R.; Griffiths, B.L.; Smith, M.J.; Foster, J.W.; Frischauf, A.M.; Lovell-Badge, R.; Goodfellow, P.N. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **1990**, *346*, 240–244. [[CrossRef](#)] [[PubMed](#)]
36. Skaletsky, H.; Kuroda-Kawaguchi, T.; Minx, P.J.; Cordum, H.S.; Hillier, L.; Brown, L.G.; Repping, S.; Pyntikova, T.; Ali, J.; Bieri, T.; et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **2003**, *423*, 825–837. [[CrossRef](#)] [[PubMed](#)]
37. Koopman, P.; Gubbay, J.; Vivian, N.; Goodfellow, P.; Lovell-Badge, R. Male development of chromosomally female mice transgenic for Sry. *Nature* **1991**, *351*, 117–121. [[CrossRef](#)] [[PubMed](#)]
38. Page, D.C.; Mosher, R.; Simpson, E.M.; Fisher, E.M.C.; Mardon, G.; Pollack, J.; McGillivray, B.; de la Chapelle, A.; Brown, L.G. The sex-determining region of the human Y chromosome encodes a finger protein. *Cell* **1987**, *27*, 67–82. [[CrossRef](#)]
39. McPherson, J.D.; Marra, M.; Hillier, L.D.; Waterston, R.H.; Chinwalla, A.; Wallis, J.; Sekhon, M.; Wylie, K.; Mardis, E.R.; Wilson, R.K.; et al. A physical map of the human genome. *Nature* **2001**, *409*, 934–941.
40. International Human Genome Sequencing Consortium Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860–921. [[CrossRef](#)] [[PubMed](#)]
41. Bernstein, B.E.; Birney, E.; Dunham, I.; Green, E.D.; Gunter, C.; Snyder, M. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74.
42. Green, P. 2x genomes—Does depth matter? *Genome Res.* **2007**, *17*, 1547–1549. [[CrossRef](#)] [[PubMed](#)]
43. Warren, W.C.; Hillier, L.W.; Marshall Graves, J.A.; Birney, E.; Ponting, C.P.; Grützner, F.; Belov, K.; Miller, W.; Clarke, L.; Chinwalla, A.T.; et al. Genome analysis of the platypus reveals unique signatures of evolution. *Nature* **2008**, *453*, 175–183. [[CrossRef](#)] [[PubMed](#)]
44. Waters, S.A.; Livernois, A.M.; Patel, H.; O’Meally, D.; Craig, J.M.; Graves, J.A.M.; Suter, C.M.; Waters, P.D. Landscape of DNA methylation on the marsupial X. *Mol. Biol. Evol.* **2018**, *35*, 431–439. [[CrossRef](#)]
45. Georges, A.; Li, Q.; Lian, J.; Meally, D.O.; Deakin, J.; Wang, Z.; Zhang, P.; Fujita, M.; Patel, H.R.; Holleley, C.E.; et al. High-coverage sequencing and annotated assembly of the genome of the Australian dragon lizard *Pogona vitticeps*. *Gigascience* **2015**, *4*, 45. [[CrossRef](#)] [[PubMed](#)]
46. Doležel, J.; Greilhuber, J.; Suda, J. Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protoc.* **2007**, *2*, 2233–2244. [[CrossRef](#)] [[PubMed](#)]
47. Duke, S.E.; Samollow, P.B.; Mauceli, E.; Lindblad-Toh, K.; Breen, M. Integrated cytogenetic BAC map of the genome of the gray, short-tailed opossum, *Monodelphis domestica*. *Chromosom. Res.* **2007**, *15*, 361–370. [[CrossRef](#)] [[PubMed](#)]
48. Mikkelsen, T.S.; Wakefield, M.J.; Aken, B.; Amemiya, C.T.; Chang, J.L.; Duke, S.; Garber, M.; Gentles, A.J.; Goodstadt, L.; Heger, A.; et al. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* **2007**, *447*, 167–177. [[CrossRef](#)] [[PubMed](#)]
49. The Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **2012**, *485*, 635–641. [[CrossRef](#)]

50. Shearer, L.A.; Anderson, L.K.; de Jong, H.; Smit, S.; Goicoechea, J.L.; Roe, B.A.; Hua, A.; Giovannoni, J.J.; Stack, S.M. Fluorescence in situ hybridization and optical mapping to correct scaffold arrangement in the tomato genome. *G3* **2014**, *4*, 1395–1405. [[CrossRef](#)]
51. Deakin, J.E.; Edwards, M.J.; Patel, H.; O’Meally, D.; Lian, J.; Stenhouse, R.; Ryan, S.; Livernois, A.M.; Azad, B.; Holleley, C.E.; et al. Anchoring genome sequence to chromosomes of the central bearded dragon (*Pogona vitticeps*) enables reconstruction of ancestral squamate macrochromosomes and identifies sequence content of the Z chromosome. *BMC Genom.* **2016**, *17*, 447. [[CrossRef](#)]
52. Collins, F.S.; Lander, E.S.; Rogers, J.; Waterson, R.H. Finishing the euchromatic sequence of the human genome. *Nature* **2004**, *431*, 931–945.
53. Miga, K.H. The promises and challenges of genomic studies of human centromeres. *Prog. Mol. Subcell. Biol.* **2017**, *56*, 285–304. [[PubMed](#)]
54. McStay, B. Nucleolar organizer regions: Genomic ‘dark matter’ requiring illumination. *Genes Dev.* **2016**, *30*, 1598–1610. [[CrossRef](#)] [[PubMed](#)]
55. Presgraves, D.C.; Balagopalan, L.; Abmayr, S.M.; Orr, H.A. Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature* **2003**, *423*, 715–719. [[CrossRef](#)] [[PubMed](#)]
56. Noor, M.A.F.; Grams, K.L.; Bertucci, L.A.; Reiland, J. Chromosomal inversions and the reproductive isolation of species. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 12084–12088. [[CrossRef](#)] [[PubMed](#)]
57. Rieseberg, L.H. Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **2001**, *16*, 351–358. [[CrossRef](#)]
58. Dumas, D.; Britton-Davidian, J. Chromosomal rearrangements and evolution of recombination: Comparison of chiasma distribution patterns in standard and Robertsonian populations of the house mouse. *Genetics* **2002**, *162*, 1355–1366. [[PubMed](#)]
59. Farré, M.; Micheletti, D.; Ruiz-Herrera, A. Recombination rates and genomic shuffling in human and chimpanzee—A new twist in the chromosomal speciation theory. *Mol. Biol. Evol.* **2013**, *30*, 853–864. [[CrossRef](#)] [[PubMed](#)]
60. Ullastres, A.; Farré, M.; Capilla, L.; Ruiz-Herrera, A. Unraveling the effect of genomic structural changes in the rhesus macaque—implications for the adaptive role of inversions. *BMC Genom.* **2014**, *15*, 530. [[CrossRef](#)] [[PubMed](#)]
61. Murphy, W.J.; Larkin, D.M.; Everts-van der Wind, A.; Bourque, G.; Tesler, G.; Auvil, L.; Beever, J.E.; Chowdhary, B.P.; Galibert, F.; Gatzke, L.; et al. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science* **2005**, *309*, 613–617. [[CrossRef](#)] [[PubMed](#)]
62. Ruiz-Herrera, A.; Castresana, J.; Robinson, T.J. Is mammalian chromosomal evolution driven by regions of genome fragility? *Genome Biol.* **2006**, *7*, R115. [[CrossRef](#)] [[PubMed](#)]
63. Larkin, D.M.; Pape, G.; Donthu, R.; Auvil, L.; Welge, M.; Lewin, H.A. Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome Res.* **2009**, *19*, 770–777. [[CrossRef](#)] [[PubMed](#)]
64. Farré, M.; Robinson, T.J.; Ruiz-Herrera, A. 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. *BioEssays* **2015**, *37*, 479–488. [[CrossRef](#)] [[PubMed](#)]
65. Capilla, L.; Sánchez-Guillén, R.A.; Farré, M.; Paytuví-Gallart, A.; Malinverni, R.; Ventura, J.; Larkin, D.M.; Ruiz-Herrera, A. Mammalian comparative genomics reveals genetic and epigenetic features associated with genome reshuffling in rodentia. *Genome Biol. Evol.* **2016**, *8*, 3703–3717. [[CrossRef](#)] [[PubMed](#)]
66. Froenicke, L.; Graphodatsky, A.; Müller, S.; Lyons, L.A.; Robinson, T.J.; Volleth, M.; Yang, F.; Wienberg, J. Are molecular cytogenetics and bioinformatics suggesting diverging models of ancestral mammalian genomes? *Genome Res.* **2006**, *16*, 306–310. [[CrossRef](#)]
67. Kim, J.; Farre, M.; Auvil, L.; Capitanu, B.; Larkin, D.M.; Ma, J.; Lewin, H.A. Reconstruction and evolutionary history of eutherian chromosomes. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E5379–E5388. [[CrossRef](#)] [[PubMed](#)]
68. Bailey, J.A.; Baertsch, R.; Kent, W.J.; Haussler, D.; Eichler, E.E. Hotspots of mammalian chromosomal evolution. *Genome Biol.* **2004**, *5*, R23. [[CrossRef](#)]
69. Batzer, M.A. The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.* **2009**, *10*, 691–703.
70. Slijepcevic, P. Telomeres and mechanisms of Robertsonian fusion. *Chromosoma* **1998**, *107*, 136–140. [[CrossRef](#)]
71. Garagna, S.; Page, J.; Fernandez-Donoso, R.; Zuccotti, M.; Searle, J.B. The Robertsonian phenomenon in the house mouse: Mutation, meiosis and speciation. *Chromosoma* **2014**, *123*, 529–544. [[CrossRef](#)]

72. Lemaitre, C.; Zaghloul, L.; Sagot, M.-F.; Gautier, C.; Arneodo, A.; Tannier, E.; Audit, B. Analysis of fine-scale mammalian evolutionary breakpoints provides new insight into their relation to genome organisation. *BMC Genom.* **2009**, *10*, 335. [[CrossRef](#)] [[PubMed](#)]
73. Van Bortle, K.; Corces, V.G. Nuclear organization and genome function. *Annu. Rev. Cell Dev. Biol.* **2012**, *28*, 163–187. [[CrossRef](#)] [[PubMed](#)]
74. Giorgetti, L.; Heard, E. Closing the loop: 3C versus DNA FISH. *Genome Biol.* **2016**, *17*, 215. [[CrossRef](#)] [[PubMed](#)]
75. Davies, J.O.J.; Oudelaar, A.M.; Higgs, D.R.; Hughes, J.R. How best to identify chromosomal interactions: A comparison of approaches. *Nat. Methods* **2017**, *14*, 125–134. [[CrossRef](#)] [[PubMed](#)]
76. Williamson, I.; Berlivet, S.; Eskeland, R.; Boyle, S.; Illingworth, R.S.; Paquette, D.; Dostie, J.; Bickmore, W.A. Spatial genome organization: Contrasting views from chromosome conformation capture and fluorescence in situ hybridization. *Genes Dev.* **2014**, *28*, 2778–2791. [[CrossRef](#)]
77. Denker, A.; De Laat, W. The second decade of 3C technologies: Detailed insights into nuclear organization. *Genes Dev.* **2016**, *30*, 1357–1382. [[CrossRef](#)]
78. Tomaszewicz, M.; Rangavittal, S.; Cechova, M.; Sanchez, C.; Fescemyer, H.W.; Harris, R.; Ye, D.; Brien, C.M.O.; Chikhi, R.; Ryder, O.A.; et al. A time- and cost-effective strategy to sequence mammalian Y Chromosomes: An application to the de novo assembly of gorilla Y. *Genome Res.* **2016**, *26*, 530–540. [[CrossRef](#)]
79. Kuderna, L.F.K.; Lizano, E.; Julià, E.; Gomez-Garrido, J.; Serres-Armero, A.; Kuhlwillm, M.; Alandes, R.A.; Alvarez-Estape, M.; Juan, D.; Simon, H.; et al. Selective single molecule sequencing and assembly of a human Y chromosome of African origin. *Nat. Commun.* **2019**, *10*, 4. [[CrossRef](#)]
80. Hughes, J.F.; Skaletsky, H.; Pyntikova, T.; Graves, T.A.; van Daalen, S.K.M.; Minx, P.J.; Fulton, R.S.; McGrath, S.D.; Locke, D.P.; Friedman, C.; et al. Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature* **2010**, *463*, 536–539. [[CrossRef](#)]
81. Cui, Z.; Liu, Y.; Wang, W.; Wang, Q.; Zhang, N.; Lin, F.; Wang, N.; Shao, C.; Dong, Z.; Li, Y.; et al. Genome editing reveals dmrt1 as an essential male sex-determining gene in Chinese tongue sole (*Cynoglossus semilaevis*). *Sci. Rep.* **2017**, *7*, 42213. [[CrossRef](#)]
82. Shao, C.; Li, Q.; Chen, S.; Zhang, P.; Lian, J.; Hu, Q.; Sun, B.; Jin, L.; Liu, S.; Wang, Z.; et al. Epigenetic modification and inheritance in sexual reversal of fish. *Genome Res.* **2014**, *24*, 604–615. [[CrossRef](#)] [[PubMed](#)]
83. Graves, J.A.M. The origin and function of the mammalian Y chromosome and Y-borne genes—An evolving understanding. *BioEssays* **1995**, *17*, 311–320. [[CrossRef](#)] [[PubMed](#)]
84. Deakin, J.E. Marsupial X chromosome inactivation: Past, present and future. *Aust. J. Zool.* **2013**, *61*, 13–23. [[CrossRef](#)]
85. Straub, T.; Becker, P.B. Dosage compensation: The beginning and end of generalization. *Nat. Rev. Genet.* **2007**, *8*, 47–57. [[CrossRef](#)] [[PubMed](#)]
86. Gu, L.; Walters, J.R. Evolution of sex chromosome dosage compensation in animals: A beautiful theory, undermined by facts and bedeviled by details. *Genome Biol. Evol.* **2017**, *9*, 2461–2476. [[CrossRef](#)] [[PubMed](#)]
87. Deakin, J.E.; Hore, T.A.; Koina, E.; Graves, J.A.M. The status of dosage compensation in the multiple X chromosomes of the platypus. *PLoS Genet.* **2008**, *4*, e1000140. [[CrossRef](#)] [[PubMed](#)]
88. Livernois, A.M.; Waters, S.A.; Deakin, J.E.; Graves, J.A.M.; Waters, P.D. Independent evolution of transcriptional inactivation on sex chromosomes in birds and mammals. *PLoS Genet.* **2013**, *9*, e1003635. [[CrossRef](#)] [[PubMed](#)]
89. Julien, P.; Brawand, D.; Soumillon, M.; Necsulea, A.; Liechti, A.; Schütz, F.; Daish, T.; Grützner, F.; Kaessmann, H. Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. *PLoS Biol.* **2012**, *10*, e1001328. [[CrossRef](#)]
90. Lemos, B.; Branco, A.T.; Hartl, D.L. Epigenetic effects of polymorphic Y chromosomes modulate chromatin components, immune response, and sexual conflict. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 15826–15831. [[CrossRef](#)]
91. Sackton, T.B.; Hartl, D.L. GBE Meta-Analysis Reveals that genes regulated by the preferentially localized to repressive chromatin. *Genome Biol. Evol.* **2013**, *5*, 255–266. [[CrossRef](#)]
92. Araripe, L.O.; Tao, Y.; Lemos, B. Interspecific Y chromosome variation is sufficient to rescue hybrid male sterility and is influenced by the grandparental origin of the chromosomes. *Heredity* **2016**, *116*, 516–522. [[CrossRef](#)] [[PubMed](#)]

93. Case, L.K.; Wall, E.H.; Dragon, J.A.; Saligrama, N.; Kremontsov, D.N.; Moussawi, M.; Zachary, J.F.; Huber, S.A.; Blankenhorn, E.P.; Teuscher, C. The Y chromosome as a regulatory element shaping immune cell transcriptomes and susceptibility to autoimmune disease. *Genome Res.* **2013**, *23*, 1474–1485. [[CrossRef](#)] [[PubMed](#)]
94. Kremontsov, D.N.; Case, L.K.; Dienz, O.; Raza, A.; Fang, Q.; Ather, J.L. Genetic variation in chromosome Y regulates susceptibility to influenza A virus infection. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 3491–3496. [[CrossRef](#)] [[PubMed](#)]
95. Yardımcı, G.G.; Noble, W.S. Software tools for visualizing Hi-C data. *Genome Biol.* **2017**, *18*, 26. [[CrossRef](#)] [[PubMed](#)]
96. Novák, P.; Neumann, P.; Macas, J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinform.* **2010**, *11*, 378. [[CrossRef](#)] [[PubMed](#)]
97. Jain, M.; Olsen, H.E.; Turner, D.J.; Stoddart, D.; Bulazel, K.V.; Paten, B.; Haussler, D.; Willard, H.; Akeson, M.; Miga, K.H. Linear assembly of a human Y centromere using nanopore long reads. *Nat. Biotechnol.* **2018**, *36*, 321–323. [[CrossRef](#)] [[PubMed](#)]
98. Tomaszewicz, M.; Medvedev, P.; Makova, K.D. Y and W Chromosome assemblies: Approaches and discoveries. *Trends Genet.* **2017**, *33*, 266–282. [[CrossRef](#)] [[PubMed](#)]
99. Dekker, J.; Belmont, A.S.; Guttman, M.; Leshyk, V.O.; Lis, J.T.; Lomvardas, S.; Mirny, L.A.; O’Shea, C.C.; Park, P.J.; Ren, B.; et al. The 4D nucleome project. *Nature* **2017**, *549*, 219–226. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).