Chapter 1

What Have Studies of Genomic Disorders Taught Us About Our Genome?

Alexandra D. Simmons, Claudia M.B. Carvalho, and James R. Lupski

Abstract

The elucidation of genomic disorders began with molecular technologies that enabled detection of genomic changes which were (a) smaller than those resolved by traditional cytogenetics (less than 5 Mb) and (b) larger than what could be determined by conventional gel electrophoresis. Methods such as pulsed field gel electrophoresis (PFGE) and fluorescent in situ hybridization (FISH) could resolve such changes but were limited to locus-specific studies. The study of genomic disorders has rapidly advanced with the development of array-based techniques. These enabled examination of the entire human genome at a higher level of resolution, thus allowing elucidation of the basis of many new disorders, mechanisms that result in genomic changes that can result in copy number variation (CNV), and most importantly, a deeper understanding of the characteristics, features, and plasticity of our genome.

In this chapter, we focus on the structural and architectural features of the genome, which can potentially result in genomic instability, delineate how mechanisms, such as NAHR, NHEJ, and FoSTeS/MMBIR lead to disease-causing rearrangements, and briefly describe the relationship between the leading methods presently used in studying genomic disorders. We end with a discussion on our new understanding about our genome including: the contribution of new mutation CNV to disease, the abundance of mosaicism, the extent of subtelomeric rearrangements, the frequency of *de novo* rearrangements associated with sporadic birth defects, the occurrence of balanced and unbalanced translocations, the increasing discovery of insertional translocations, the exploration of complex rearrangements and exonic CNVs. In the postgenomic era, our understanding of the genome has advanced very rapidly as the level of technical resolution has become higher. This leads to a greater understanding of the effects of rearrangements present both in healthy subjects and individuals with clinically relevant phenotypes.

Key words: Copy number variation, Recombination, Gene dosage, Mosaicism, Insertional translocation, Comparative genomic hybridization

1. Introduction

In the last century, researchers viewed our genome as an imperturbable collection of base pairs in which small mistakes (i.e., single base pair changes) were the predominant cause of inherited disease. Today, we have a broader understanding of the plasticity of our genome.

Lars Feuk (ed.), *Genomic Structural Variants: Methods and Protocols*, Methods in Molecular Biology, vol. 838,
DOI 10.1007/978-1-61779-507-7_1, © Springer Science+Business Media, LLC 2012

This will be made evident by the numerous examples of genomic alterations described in this book. Depending on the segments (and genes) involved one may encounter different results: if deviations from the normal diploid state convey a phenotype, these conditions are referred to as genomic disorders (1, 2). In other cases, segments that either do not contain genes or that contain genes or regulatory regions which are not dosage-sensitive, may be involved in rearrangements. These alterations have been associated with both small-scale changes reflected in individual genome variation (a few base pairs to a few Kb), and large-scale genomic changes that are obvious throughout evolution of primate and hominid genomes (3–7).

The first reports of diseases caused by mega base sized genomic rearrangements were published in the early 1990s (8). Thanks to technological and conceptual advancements, hundreds of distinct disorders, and thousands of copy number variant (CNV) (i.e., deviating from the normal diploid state; usually $n=2$) regions have now been described.

Genomic rearrangements include changes in the diploid genome that lead to duplication, deletion, insertion, inversion, or translocation of segments. Such changes can range from a few hundred base pairs to mega bases and may lead to neutral polymorphisms or disease phenotypes (9). It is also known that CNVs are ubiquitous, involving up to 12% of the human genome (10–20) and may arise during meiosis (21) or mitosis, as apparent from somatic CNV mosaicism studies (22–24).

The Human Genome Project generated the first reference haploid genome, which, together with the development of high-resolution genome analysis techniques, like array-based methods, and the refinement of those methods have enabled total genome-wide analyses to rapidly proceed. In less than a century, the substance of heredity was identified; its structure elucidated, the genetic code deciphered, the genome sequenced, and corresponding base pairs accurately described. Despite the overwhelming (and constantly growing) amounts of data, there are still many questions that remain unsolved.

Structural rearrangements of the human genome are of two general types: recurrent and nonrecurrent. Rearrangements with recurrent end-points show breakpoint clustering and junctions that are limited to the location of low copy repeats (LCRs) (described below) where homology is extensive. In this case, the LCRs both stimulate and mediate the rearrangement by acting as homologous recombination (HR) substrates. Nonrecurrent structural changes, on the other hand, have scattered breakpoints and the boundaries share limited or no nucleotide identity (i.e., homology). Here, the resulting rearrangement can be complex and tends to have one breakpoint that groups close to highly polymorphic LCR rich regions. In this case, it is possible that the presence of LCRs may stimulate the formation of secondary DNA structural conformations that can lead to genomic instability. Different LCR

conformations can provide single-stranded regions that may result in collapsed DNA replication forks. Such events can generate one-ended, double-stranded DNA breaks or conformations that can result in two-ended double-stranded breaks and must be repaired by classical double-strand break repair (DSBR) (25).

The focus of this chapter is to illustrate *which mechanisms* lead to genomic rearrangements, the genomic disorders that result from such rearrangements and the *knowledge acquired* through such genomic studies.

2. Architectural Features of the Human Genome

A little over half our genome is made up of both *repeat sequences* and *highly repetitive elements*. The main difference between these two is primarily the frequency, with the former presenting very few copies (i.e., LCRs) that have arisen as segmental duplications of the original sequence and seem to make up independent sets (i.e., they originate from many different sequences each copied just a few times), and the latter appear to have arisen from duplication events of the same sequence (or same type of sequences) that has occurred several thousand times (see Table 1).

Table 1
Classification and approximate percentage of repeat sequences in the Human genome (International Human Genome Sequencing Consortium, Nature, 2001, 2004, Levy/Diploid Genome PloSB/2007)

Category	Type	Length of one unit	Copy number (per genome)	Approximate percentage of the human genome	
Transposon-derived repeats (Interspersed repeats)	LINEs	6 kb	850,000	21	45
	SINEs	100–400 bp	1,500,000	13	
	LTR retrotransposons	1.5–11 kb	450,000	8	
	DNA transposons	80 bp–3 kb	300,000	3	
Simple sequence repeats	Direct repeats of short *k*-mers (minisatellites, microsatellites)	1–500 bp	?	4.4	
Low copy repeats or segmental duplications	Blocks of DNA that have been copied from one region of the genome into another region	1–300 kb	?	5.3	
Blocks of tandemly repeated sequences	Centromeres, telomeres, short arms of acrocentric chromosomes and ribosomal gene clusters	?	?	8	

Sedimentation equilibrium centrifugation data provided the first line of evidence that demonstrated the presence of repeat DNA in an experimental fashion. Researchers found that for most eukaryotes DNA was divided into a main band (or peak) and any additional peaks were dubbed "satellite DNA." A few years later, reassociation kinetics assays proved that these peaks were made up of highly repetitive DNA (reviewed in (26)). Today, the term "satellite DNA" depicts tandemly repeated sequences.

Table 1 shows four categories of repeat DNA: Interspersed repeats, simple sequence repeats, low copy repeats and blocks of tandemly repeated sequences. *Interspersed repeats* are by far the most abundant, among them *LINE1* (LINE) and *Alu* (SINE) elements are prominent as these have 850,000 and 1,500,000 copies per haploid genome, and they comprise 18.9 and 10.6% of our genome, respectively. Among the *simple sequence repeats* (or microsatellites), dinucleotides are the most abundant and make up 0.5% of our genome, half of that being specifically AC repeats. These also include minisatellites (also called VNTRs, variable number tandem repeats) that are highly polymorphic. There is less information on *tandemly repeated sequences*, like centromeres, telomeres, pericentromeric and telomeric regions, as these have been purposefully underrepresented in sequencing projects mostly because of the difficulty to find the correct genomic location of clones containing such repetitive sequences (for more detailed info, see ref. 27).

LCRs, (28) or segmental duplications (SDs, (29)) are defined as DNA segments that occur more than twice in the haploid genome, have an extension larger than 1 kb (may extend to over 300 kb in size) and present more than 95% sequence identity between their paralogous copies (4, 30). LCRs are known to be one of the mediators of a large number of genomic rearrangements, however, all repetitive DNA can potentially be a substrate for rearrangements. The presence of these specific repetitive sequences and repeats as architectural features of our genome has been shown to convey genomic instability.

Some genomic loci have a complex genomic architecture, meaning that they bear complicated patterns of direct and inverted LCRs. These regions are associated with what are called CNV regions where the copy number can vary from the expected diploid genome number, usually $n=2$, one inherited from each parent (i.e., one copy of each segment per haploid genome) (25).

Genomic changes can occur at three distinct levels: (a) at the base pair level (SNPs, point mutations), (b) at the structural level (chromosomal rearrangements, genomic disorders) or (c) at the conformational level. Bacolla and Wells reviewed how specific DNA sequences can lead to conformational changes of DNA

and the formation of non-B DNA structures. For example: inverted repeats, also called mirror repeats, form cruciforms; purine-pyrimidine tracts can form triplex DNA; alternating purine–pyrimidine tracts can form left handed Z-DNA; and four runs of closely spaced guanine seem to be able to form tetraplex DNA (31). A series of experiments using plasmids in which the LacZ-GFP system was incorporated as a reporter adjacent to a poly purine–pyrimidine tract (containing direct and inverted repeats) increased the observed frequency of mutational events. This was evident by the disruption of the GFP gene. Intra- and intermolecular recombination was observed. In most cases it lead to deletion, and in a particular instance, to an inversion. In all deletions examined, the breakpoints occurred at the predicted non-B DNA structures and left a microhomology *scar* (2–8 bp) at the breakpoint site. The authors suggest that non-B DNA conformations increase the frequency of appearance of gross rearrangements by increasing the number of double-strand breaks (DSBs): their data shows that a primary DSB in one site seems to trigger the appearance of a second DSB at a secondary site on the same or different DNA molecule to produce a recombination reaction between these two DSBs (32, 33).

Evidence for this is also provided in work done with yeast (34) and mammalian cells (35), where artificial induction of DSBs can lead to reciprocal translocation by the non-homologous end joining (NHEJ) pathway (explained below). Using chromosome conformation capture (3C) it has been shown that gross chromosomal rearrangements can be generated after the formation of DSBs in yeast chromosomes in which DSBs become sequestered by the telomerase machinery and the Mps3p nuclear envelope protein to the nuclear periphery, which limits the rate of HR. The results suggest that there might be a competition between alternative repair mechanisms and that the nuclear location might affect the outcome of the repair by modifying the interaction of the DSB with other chromatin fragments (36).

CNVs and structural change seem to cluster in regions were LCRs are present, in heterochromatic regions (centromeres and telomeres) (37–41), in replication origins and terminators (42), where scaffold attachment sequences are present (43, 44) and where LINE and SINE elements are prevalent (45).

In addition to the chromosomal architecture, the type of cell division (mitosis, meiosis), the stage of the cell cycle and the characteristics of the broken segment (two-ended vs. one-ended double-stranded DNA break) may lead to the recruitment of different factors that utilize specific types of repair mechanism, leading to genomic rearrangements.

3. Mechanisms Leading to Rearrangements Observed in Disease

There are a number of proposed mechanisms that lead to rearrangements of our genome that can generate CNV. These mechanisms are commonly classified as HR mechanisms and nonhomologous recombination mechanisms, the latter are further subdivided into replicative and nonreplicative mechanisms (reviewed in (25)). Our focus is on mechanisms that lead to rearrangements associated with disease, so we concentrate on *non-allelic homologous recombination* (NAHR), NHEJ, and *fork stalling and template switching* (FoSTeS)/*microhomology-mediated break-induced replication* (MMBIR).

3.1. NAHR

NAHR is a homologous recombination mechanism. HR is a type of DNA repair mechanism with other functions in dividing cells, like permitting ordered segregation of chromosomes and producing new combinations of linked alleles (meiosis). HR in mammalian and human cells (46, 47) requires around 200–300 bp of near identical sequence, a minimal efficient processing segment, and the participation of the Rad51 protein, which catalyzes the invasion of a 3′ end of ssDNA to a duplex sequence either on the sister chromatid or the homolog. HR is considered an accurate repair mechanism, but may lead to structural rearrangements because multiple paralogous LCR tracts may be present in contiguous regions, thus *confounding* the repair machinery. During repair, the utilization of paralogous LCRs as homology substrate sequences leads to NAHR (25).

The mechanism of NAHR as a cause of rearrangements associated with genomic disorders was first described in the early 1990s (48), but the term NAHR was not introduced until 2002 (30). Also know as ectopic homologous recombination, NAHR can cause a variety of rearrangements. This includes duplications, deletions and inversions, which may take place between LCRs on the same or different chromosomes, in direct or opposite orientation. Deletion or duplication can result when two LCRs are positioned in the genome in direct orientation allowing interchromatidal or interchromosomal exchanges. When the interacting LCRs cause intrachromatidal NAHR, only deletions are observed. If the LCRs involved in the rearrangement are on the same chromosome but in opposite orientation, NAHR will result in the inversion of the segment contained between LCR substrate copies. Translocations can result if the LCRs mediating the rearrangement are located on different chromosomes (1). Based on this information, we expect the frequency of deletions to be higher than the frequency of duplications, and single sperm PCR assays seem to support this hypothesis, at least for the loci examined (21, 49, 50).

Repetitive elements, such as SINEs, LINEs, and LTRs also seem to be able to mediate NAHR (11, 51, 52) even though the

identity stretch between any two of these elements is usually shorter than what has been described for LCRs.

When NAHR is operative, the crossovers cluster in narrow hotspots (53). This has been observed in diseases like Neurofibromatosis Type 1 (NF1, MIM162200), Charcot-Marie-Tooth disease Type 1A (CMT1A, MIM118220) and Hereditary Neuropathy with Liability to Pressure Palsies (HNPP, MIM162500).

NAHR may occur during meiosis or mitosis, the first usually responsible for inherited disorders (like CMT1 and HNPP) and sporadic disorders (like PTLS or SMS), the latter responsible for somatic mosaicism, which may lead to cancerous tumor formation (53–55). There does not seem to be a distinction between the sequences involved in meiotic or mitotic NAHR (for examples, see ref. 54–56), but the actual recombination hotspots and the frequency of mitotic or meiotic recombination can differ (53). For example, a study on NF1 shows that mitotic and meiotic rearrangements seem to be responsible for different deletion sizes (i.e., uses different recombination substrates) and suggest different mechanisms may be operating in each case. NF1 can occur as a consequence of intragenic mutations or microdeletions involving the *NF1* gene. The most common deletion (type-1) spans 1.4 Mb on 17q11.2 and encompasses 14 genes, including *NF1*. Type-1 deletions predominantly involve the maternal chromosome by interchromosomal recombination during meiosis. Type-2 microdeletions spans 1.2 Mb, and hence the patients present a less severe clinical phenotype. Both type-1 and type-2 microdeletions seem to be mediated by NAHR. In addition, patients with atypical microdeletions have also been identified. Atypical NF1 deletions seem to arise as a consequence of a nonhomology-based mechanisms (like NHEJ) and take place on the paternal chromosome by an intrachromosomal mechanism during mitotic cell division in spermatogenesis (57).

The efficiency of NAHR may be affected by several factors including: (a) degree of sequence identity, (b) orientation of the LCRs, (c) distance between them, (d) their location (intra- or interchromosomal) provided that the recombination event occurs during mitosis or meiosis, and (e) on the sex of the individual (oogenesis or spermatogenesis) (53).

NAHR is distinguished from unequal crossing-over since the latter refers to the segregation of marker genotypes and the phenomena observed (recombinant chromosomes). NAHR is mechanistic; its products include inversions, and because of the architectural features involved, allow us to make specific outcome predictions depending on the HR substrate orientation.

When PTLS-associated uncommon recurrent duplications were investigated using aCGH and recombination hot spot analyses, it was found that the crossover occurred close to, within 400 nucleotides, a recently described homologous recombination hotspot motif. This *cis*-acting sequence appears to bind PRDM9,

a protein with histone H3K4 trimethylase activity. The motif was defined during HapMap studies as an allelic homologous recombination (AHR) stimulating sequence associated with recombination "hotspots." However, it is found nearby both AHR and NAHR "hotspots" for crossovers (58).

3.2. NHEJ

NHEJ is a non-HR, nonreplicative repair mechanism. In mammalian and yeast cells, NHEJ is one of the primary repair mechanisms used to resolve DNA DSBs and seems to function in all phases of the cell cycle, but especially in G1 phase. DSB can occur because of the presence of reactive oxygen species, that may arise as a consequence of endogenous (by-products of cellular metabolism) or exogenous phenomena (X-rays, gamma-rays). Not all DSBs are stochastic events; some are actually programmed in cells as is the case for V(D)J recombination to generate antibody, T cell receptor diversity (mitotic) (59) and crossing-over (meiotic).

Mammals possess two mechanisms to repair DSBs: homologous repair (HR, described above) and NHEJ. NHEJ requires the binding of a Ku complex (specifically, Ku70 and Ku80, (25)), which recognizes the break, and the participation of numerous protein complexes. The most studied complexes are (a) DNA-dependent protein kinase (DNA-PK) apparently involved in the tethering of broken ends to facilitate rejoining and recruiting/activating proteins responsible for chromatin remodeling, DNA end-processing and ligation; (b) DNA Ligase IV-XRCC4 complex, which is present in all eukaryotes, stimulates DNA ligation; and (c) The participation of one or more end-processing enzymes (mainly exonucleases).

NHEJ is considered an error-prone repair mechanism since it does not rely on the presence of a homologous template. Unlike NAHR, NHEJ does not require homologous sequence substrates nor minimal efficient processing segment (MEPS), and as a consequence, small deletions (1–4 bp) may be apparent or a few nucleotides (2–34 bp) of free DNA (usually of mitochondrial or retrotransposon origin) can be added to the broken ends and remain in the junctional sequence (60–65).

While NHEJ does not depend on the presence of LCRs, the occurrence of repetitive DNA elements and sequences related to architectural modification of DNA seem to cause genome instability and susceptibility to DSBs which may be repaired by NHEJ. Examples of genomic disorders that can occur by NHEJ are some nonrecurrent rearrangements associated with Pelizaeus–Merzbacher disease (PMD, MIM312080, (66)) and Smith–Magenis syndrome (SMS, MIM182290, (51)).

3.3. FoSTeS/MMBIR

Studies of *E. coli* show that replication inhibition leads to the formation of DSBs (67). Normal human cells that are induced to

replicative stress by using aphidicolin, a drug that is able to inhibit DNA polymerases associated with replication, show numerous, large copy number changes that have microhomology at the join points. Such occurrences could be explained by erroneous repair happening after the replication fork has stalled (68).

In many cases, genomic disorders seem to be caused by complex rearrangements in which duplicated and triplicated segments are interrupted with regions of no observable copy number change. Such is the case for several genes, like *PLP1*, *MECP2 APP*, *SNCA*, *RAI1*, *PMP22* (69–82), which show duplications and triplications that are causal of the disease and associated phenotypes. In 2007, Lee et al. (83) proposed a replication-based mechanism that could lead to such complex human genomic rearrangements: FoSTeS. The authors studied patients with PMD previously determined to have a duplication of the dosage-sensitive gene *PLP1* using a high-resolution oligonucleotide array (comparative genomic hybridization assay) and breakpoint sequence analysis. Many patients, in fact, had complex rearrangements, in which duplicated and triplicated segments were interspersed with normal-copy-number sequences. They sequenced breakpoints in 3 out of 17 patients and described extremely complex rearrangements, which could not be parsimoniously explained by either NAHR or NHEJ. FoSTeS/MMBIR is proposed to occur during mitosis. It also represents a type of non-HR mechanism since little or no homology is necessary.

According to the FoSTeS model, a nick (or ssDNA lesion) could lead to stalling of the replisome at the fork, and the lagging strand would disengage and switch to another active replication fork. Microhomology would be necessary for the priming of DNA replication on the switched template, and the direction of the newly inserted segment is dependent of the direction in which the replisome is advancing (5′ to 3′ or 3′ to 5′) (83).

The signature experimental observations described by Lee et al. (83) (i.e., complex rearrangements with microhomology at the breakpoint junctions) are similar to those described in *E. coli*, whereby starvation stress induces amplification of the *lac operon* to 20–100 copies that appear both in direct and inverted orientation (84–86). In this system, as shown by evidence evaluating the effects of the deletion or over-expression of a 3′ exonuclease gene (*xonA*), free 3′ DNA ends seem to be involved in the amplification of the *lac operon*. The fact that duplication is more prevalent in strains where *xonA* is deleted and similar experiments with a 5′ exonuclease show no relevant change in the amplification number, support the notion that 3′ ends are required for this mechanism (87).

Lac system assay in *E. coli* provided for an amplification model in which replication is restarted at sites of DSB repair of DNA, where template switching (to a different replication fork) was hypothesized to occur during replication restart at stalled replication forks (87). Also, double-strand cleavage of DNA close to the operon augments amplification rate (88).

Studies in yeast seem to support the hypothesis that repeats may arise in the genome as a consequence of repair during DNA replication (42). In this experiment, they used a topoisomerase I inhibitor to nick DNA. These nicked strands caused fork collapse, and ultimately increased the frequency of duplication formation. Interestingly, the authors also observed that fork stalling, as opposed to fork collapse, did not have a strong impact on the frequency of duplication (42). Experiments in which DNA was transfected into mammalian cells revealed microhomology and erratic sequence insertion at the junctions, which the authors explained by proposing a similar model also based on template switching (89).

MMBIR was recently proposed by Hastings et al. (84) to account for the microhomology observed in complex rearrangements, template switching, and the insertion of short sequences at the breakpoint junction that are templated from nearby genomic intervals. They suggest that a one-ended, double-stranded DNA molecules generated by a collapsed replication fork, from stalled transcription complexes, at excision repair tracts, or at secondary structures in DNA (i.e., cruciforms, hairpins), can be repaired by using available single-stranded DNA (also available in secondary DNA structures) if it shares very short homology with 3′ end of the broken strand that has escaped from the collapsed fork. At the molecular level, they propose that a stress-induced reduction of Rad51 leaves repair of the collapsed replication fork in the hands of Rad52 (and possibly other proteins), which requires only minimal homology and would use annealing of single DNA strands to prime DNA replication as the main mechanism available for such repair. In this model, MMBIR is substituting classical break-induced replication (BIR). The MMBIR model can explain complex rearrangements and can also explain the formation of simple duplications, deletions, inversions, translocations, and amplifications due to rolling circles depending on the location that the lagging strand switches to: a position behind the location where the fork collapsed, to a nonhomologous sequence, etc. The authors suggest numerous implications that support a replication repair model, starting with cancer formation and genomic disorders and moving all the way up to exon shuffling and evolution. MMBIR is a molecular mechanistic model based on studies in both human and model organisms, such as *E. coli* and yeast. The FoSTeS model was based upon the "phenomenology" observed in human genomic disorders; it does not provide the mechanistic detail as elucidated for MMBIR.

4. High-Resolution Genome Analyses Methods and Their Limitations

Initially, the only way to visualize our entire genome was by performing karyotypes and studying chromosomes. Some human disease phenotypes were found to be the consequence of whole chromosome aneuploidy or segmental aneuploidy, the presence of large deletions that spanned over 5 Mb in length. Fluorescence in situ hybridization (FISH) was the next approach used to identify and delimit duplications and deletions (90). In most cases, BAC clones (150 to 200 Kb) were used as probes, providing a resolution of deletions under 1 Mb in size (53). Fosmid clones (~40 kb) were also used (57).

Currently, chromosomal banding and FISH are still performed, but CNV, for both research and clinical purposes is mostly studied in diagnostic laboratories by array-based copy number analysis (ABCNA) due to its superior level of genome resolution. Two types of platforms are widely available to perform ABCNA: *array comparative genomic hybridization* (aCGH), in which the test sample is directly compared to a gender matched control sample (91), and the *noncomparative arrays* (NCA) that determine the relative copy number within a single genome, in a quantitative manner, without the use of a control sample in the same experiment.

An array consists of a collection of DNA fragments (BACs, PACs, cDNA, PCR products or synthetic oligonucleotides) that are attached to a glass or silica slide (92). The two leading companies that produce oligo CGH arrays are Agilent and NimbleGen (Roche). By labeling the patient and control DNA with different fluorescent dyes, one can evaluate copy number changes depending on the color and intensity of the signal (green vs. red) recorded for each interrogating probe on the array. Agilent allows for custom designed arrays that may span a segment of interest or the full genome. They offer single or multipack formats (1, 2, 4, or 8 arrays per slide) and up to 1 million interrogating 60-mer probes can be *printed* onto each piece of glass. It is possible to produce custom or predesigned arrays (http://www.chem.agilent.com). NimbleGen HD2 arrays also use long oligo probes (50–75 bases in length) with a maximum number of 2.1 million probes per array, making it the currently available highest density CGH. Their arrays can be ordered in multiplex formats (1, 3, or 12 arrays per slide). Custom-designed arrays can also be produced (http://www.nimblegen.com).

Advantages of using oligo CGH arrays are numerous, some of the most obvious being (a) dividing cells are not required and small amounts of DNA can be used (350 ng to 2 μg, depending on the format), (b) it detects unbalanced chromosomal abnormalities at a level that escapes banded karyotype resolution, (c) aCGH detects genomic mosaicism previously not detected by karyotype analysis, and (d) both LCRs and repetitive sequences may be excluded.

The main disadvantage of this technology is that it is a comparative methodology, in which the result is relative, instead of absolute. The final signal depends on the copy number status of the control DNA; for example, what looks like a gain on patient DNA, might actually be a loss in the control sample (93). Other disadvantages of aCGH include (a) the inability to provide positional and orientational information, which is reflected in the failure to detect balanced inversions and balanced translocations (93), and (b) the difficulty of identifying the total copy number of genes that have more than four copies. This last disadvantage represents a technical limitation of the method that can be explained as an overall decrease in dynamic range as the comparative relative copy number increases. This will certainly be one of the next challenges of copy number identification: dealing with disease states that result from copy numbers greater than 4 (and in most cases, even greater than 3). There are several genes that are present in multiple copies like immune function genes. Once this issue resolves the evaluation of some of complex traits like lupus and autoimmune disorders will become more feasible.

Recently, next-generation sequencing platforms (NGS) were used to develop a read-mapping algorithm (mrFAST) that allows assessment of CNV and accurately predicts the absolute copy number of multicopy genes and genomic regions (94). It is quite possible that the next level of genomic resolution might be whole genome sequencing and an even greater number of CNVs may be found: with the change in technology, the number of described CNVs seems to increase at least an order of magnitude (see Table 2).

NCAs are primarily produced by Affimetrix and Illumina. Affimetrix offers a hybrid genotyping array (Genome-Wide Human SNP Array 6.0) that places 906,600 SNPs and 946,000 nonpolymorphic probes on the same array. The probes are evenly spaced

Table 2
Genome resolution and copy number variation

Interrogating probes (pixels)	Total CNVs detected	Average number of CNVs in an individual	References
ROMA 85 K	221	11	(12)
BAC aCGH 3 K	255	12	(10)
BAC aCGH 25 K Affymetrix GeneChip Human Mapping 500 K	1,447	~24–70 depending on the platform	(13)
Oligonucleotide aCGH 42×10^3 K	11,700	~1,117–1,488	(138)

aCGH array-based comparative genomic hybridization, *BAC* bacterial artificial chromosome, *ROMA* representational oligonucleotide microarray analysis

and target known genomic regions that exhibit CNV. The test sample is processed, labeled, and directly hybridized to the array, without the use of a control sample. Their software uses median absolute pairwise difference (MAPD) values to predict copy number.

Illumina offers Infinium HD BeadChips that provide from 300,000 to over a million markers. Illumina arrays are different from the other available commercial arrays in that the oligonucleotides are placed on 3-micron silica beads that self assemble in microwells on fiber optic bundles or planar silica slides (BeadArray Technology). The GenomeStudio Software allows for CNV data analysis by using the cnvPartition algorithm, which computes the output values of the log*R* ratio (LRR) and B allele frequency (BAF). Since LRR is the log ratio of observed probe intensity vs. expected intensity, deviations from zero are interpreted as a copy number change.

Both Affimetrix and Illumina offer options for developing custom arrays although these platforms are not as flexible as the aCGH format for custom arrays. The advantages/disadvantages of these platforms are very similar to those described for oligo CGH arrays, except that they are not a comparative assay. Advantages over aCGH include the ability to: identify consanguinity relationships, infer loss (LOH) or absence (AOH) of heterozygosity and differentiate alleles and parental origin using the SNP information.

Being able to analyze our genome at this new level of resolution has permitted a more thorough study of normal and deleterious copy number changes and rearrangements. It has also allowed us to delve into the mechanisms resulting in rearrangements and even to propose new mechanisms that may lead to genomic rearrangements. Such was the case for FoSTeS (53, 83) and MMBIR (84).

5. What Have We Learned from High-Resolution Analyses of the Human Genome?

The increase in the level of resolution at which we can examine the human genome has been proportional to the gain in knowledge relative to the structure, fluidity, and mechanisms leading to changes in our genome, both normal and related to disease. Some of the most unanticipated but interesting findings are described below (see Fig. 1).

5.1. Mosaicism

Chromosomal mosaicism is defined as having more than one cell line with distinct karyotypes in different cells of one individual. When karyotypes and chromosomal banding are used for this type of screening, the identification of mosaicism is limited to the cell type used and percent required for detection. Conventional chromosome studies use stimulated peripheral blood cultures that

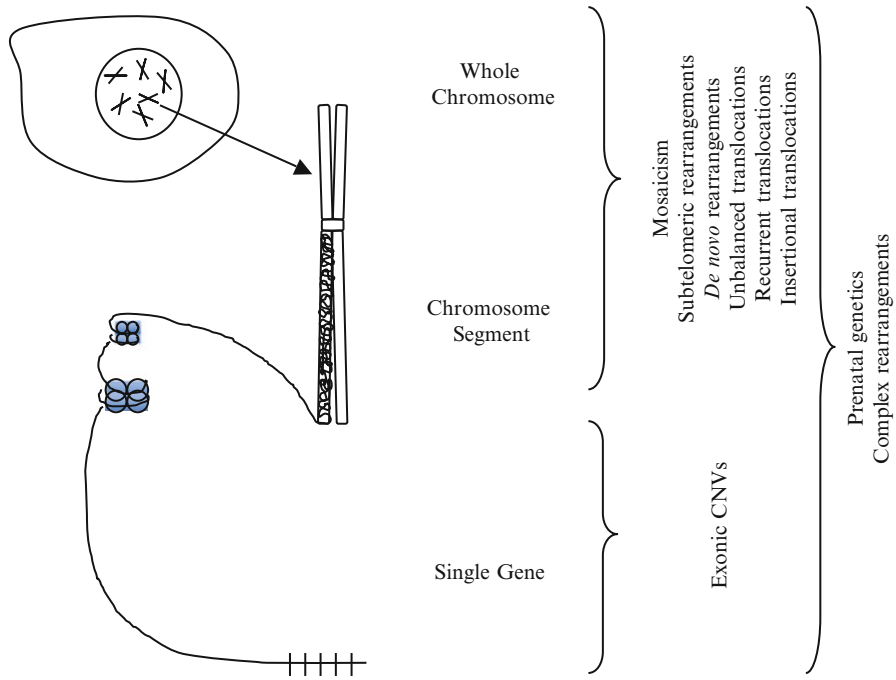

Fig. 1. What studies of genomic disorders have taught us about our genome: Findings according to the size of segments involved in the rearrangements.

depend on the use of artificially added mitogens, a selection step that limits the use of conventional chromosome analysis in mosaicism studies.

Targeted BAC clone arrays were used to screen 2,585 clinical samples and reported finding 12 (0.46%) that presented chromosomal mosaicism (95). Of those, 10 (0.39%) patients were previously reported to have a normal blood chromosome analysis. Also using array CGH, (96) reported 18 cases where mosaicism was detected, 14 of which (8% of all abnormal cases evaluated) were previously unknown. Lu et al. reported chromosomal mosaicism in 1.9% of neonates with birth defects (97).

In a more recent study, mosaicism sensitivity was measured by using artificially derived whole chromosome and segmental aneuploidies. Surprisingly, the authors found that oligonucleotide aCGH can detect low-level mosaicism in both cases, about 10% mosaicism for whole chromosome and 20–30% mosaicism of segmental aneuploidies (98). The higher mosaicism detection rates seem to be due to the fact that (a) DNA is extracted from all white blood cell lineages to perform microarray analysis, (b) cell "population" analysis of arrays versus the single cell events studied by karyotypes, and (c) the greater ability to detect subtle copy number changes with arrays; making oligonucleotide aCGH a powerful tool for mosaicism detection.

Cumulative evidence supports the notion that many of us might be mosaics due to somatic rearrangements occurring in different tissues. Studies of monozygotic twins with either concordant or discordant clinical phenotypes showed the presence of unlike CNVs within twin-pairs in both groups (22). In the past, we have assumed that normal cells are genetically identical and that most CNV must be generated during meiosis. One study analyzed CNVs in 34 tissue samples (brain, skin, heart, kidney, smooth muscle, liver, etc.) from three subjects and found CNVs that ranged in size from 82 to 176 kb, which affected a single organ, or one or more tissues of the same subject (99). These data indicate that humans might regularly present somatic mosaicism and suggests the involvement of somatic CNVs in tissue-specific disorders, such as cancer. In a study on mouse embryonic stem cells (23), extensive and recurrent CNVs generated in the clonal isolates derived from common parental lines were observed. They proposed that novel CNVs likely arose during mitosis and that all somatic tissues in individuals may show mosaicism in the form of variants of the zygotic genome. A study on chromosomal inversions in human DNA derived from blood (24) reported recurrent genomic inversions were found at a relatively high frequency in blood cells using long-range PCR assays. The rearrangements were more abundant in adults than in newborns, concluding that cell populations should be considered as mosaics with regard to their genomic structure and that mosaicism increases with age. The case for antibody diversity occurring as a consequence of programmed rearrangements (59) is an example of how mitotic rearrangements in specific cell types may represent developmental biology pathways or systems that utilizes genomic rearrangements to implement a specific cell-type programming repertoire.

5.2. Subtelomeric Rearrangements

Developmental delay, mental retardation, dysmorphic features, and congenital anomalies may be caused by subtelomeric rearrangements. Subtelomeric regions are gene-rich and because of the presence of repeat sequences are susceptible to genomic rearrangements (100, 101). Traditional methods to assess subtelomeric rearrangements include karyotype, chromosome banding and subtelomeric FISH. The use of aCGH or chromosomal microarray analysis (CMA) with extended subtelomeric coverage has enhanced the detection of rearrangements in these genomic regions.

Targeted aCGH with extended coverage at subtelomeric regions, up to 10 Mb of the 41 subtelomeric regions, was first introduced at Baylor College of Medicine in 2004 as a clinical test coupled with FISH verification (102). In one study, CMA was applied to assess 5,380 clinical patients and identified 499 (9.3%) cases with subtelomeric imbalances, of which 236 (4.4%) were pathogenic. This represents close to half of all genomic abnormalities detected using those specific arrays (37), and was significantly

higher than that reported by other groups using other technologies with less genome resolving capability. Approximately 2.5% pathogenic subtelomeric rearrangements were detected by subtelomeric FISH (103, 104).

A high proportion of cytogenetic abnormalities are due to rearrangements of subtelomeric regions and aCGH is a sensitive and robust platform that can be used to detect them (37, 105–108).

Subtelomeric rearrangements of 9q34.3 were never identified prior to subtelomere FISH as they were too small to be seen by conventional microscopy. Analysis of 43 breakpoints within the 9q34 region of patients that presented subtelomeric deletions using array-CGH showed that short repetitive elements, such as SINE, LINE, LTRs, and STRs, were present at or near the breakpoints. Such elements are susceptible both to DSBs, because of the formation of secondary structures, and also to accumulation of single-strand breaks in the replication fork. The presence of these short repetitive elements in subtelomeric regions has been proposed to aid in the stabilization of terminal deletions. In one patient, P6, the presence of an interrupted deletion/inverted duplication structure is proposed to be the scar of a breakage-fusion-bridge cycle mechanisms. However, other complex structures can be more parsimoniously explained by a FoSTeS/MMBIR mechanism than by multiple DSBs healed by NHEJ (38).

5.3. De Novo Genomic Rearrangements and Sporadic Birth Defects

Conventional diagnostic genetic analyses usually perform single locus (gene) testing by DNA sequencing with the expectation to find a point mutation that will be causal for a specific phenotype. It has been described that copy number changes can be a common cause of genetic disorders, where *de novo* locus-specific mutation rates for genomic rearrangements are between 10^{-6} and 10^{-4}, 100- to 10,000-fold greater than that of point mutations (110).

According to the 2008 report generated by the National Center for Health Statistics (http://www.cdc.gov/nchs/hus.htm), the leading cause of death in the neonatal period (less than 28 days of life) are disorders related to short gestation, low birth weight and congenital malformations. The increase in genome resolution has provided clinical geneticists a finer tool, enabling them to identify a causal CNV in patients in which the presence of a chromosomal abnormality is suspected. When using traditional cytogenetic analysis, structural rearrangements have been detected in 0.5% of newborns (110). When a physician suspects a chromosome syndrome, the detection rate of a chromosome abnormality is about 21% (111–113).

In a study published in 2008 (97), a total of 638 neonates with different birth defects were evaluated using CMA. This is one of the largest published studies that used aCGH to estimate the frequencies of genomic imbalances testing neonates with birth defects. The results are astonishing. The reported pathological CNV detection

rate for subject samples referred with an indication of "suspected chromosomal abnormality" was 66.7%. When compared to other studies published during the previous 15 years (111–114), this number is about three times greater (66.7 vs. 21.6%). It is clear that these rearrangements could not have been identified nor defined by GTG-banded karyotype and illustrates the importance in resolution of the genome for diagnostics.

5.4. Prenatal Genetics

Balanced de novo chromosomal rearrangements have been observed in about 6% of prenatal evaluations (115) and are commonly found in both patients with phenotypic abnormalities and in seemingly normal individuals. Of these, ostensibly 1 out of every 5 is a de novo event (110); *de novo* apparently balanced translocations have a greater chance of being associated with an abnormal clinical outcome (115).

In 2006, an analysis of 30 uncultured, previously characterized prenatal samples using two different BAC and PAC arrays to detect prenatal chromosome abnormalities was performed (105). Two types of arrays were used: a "large" array covered the entire genome at a 1-Mb resolution, and a "small" array, specifically targeted regions of clinical interest. The use of the small array allowed the correct diagnosis of 29/30 samples (the exception was a case of triploidy). The authors concluded that aCGH could potentially replace conventional cytogenetics for the majority of prenatal diagnosis, but warned that using large arrays could generate difficulties in interpretation until more was learned about genomic CNV. Aggregate data has shown in recent years that all rearrangements do not lead to disease, and many appear to represent polymorphic variants in the population (see Database of Genomic Variants, http://projects.tcag.ca/variation/). Similar findings were reported by (106).

The practicability of using targeted array CGH for genomic imbalance assessment of current pregnancies was evaluated (116). Ninety-eight samples of amniotic fluid, chorionic villi, and cultured cell were analyzed, and they all showed complete concordance between karyotype and array results. The authors demonstrated the feasibility of using aCGH for prenatal diagnosis and suggested that the use of arrays could increase the detection of abnormalities relative to risk. They also reported a shorter turnaround time (6–16 days). Chromosome banding and microscopic karyotyping is a labor intensive technique that is not amenable to automation and, because of the need to culture cells, has an average turnaround time of 2 weeks.

Restricting the analysis of a prenatal sample to only kayrotype or G-banding may limit the information available for counseling and informed decision making on behalf of the parents. This was nicely shown using G-banded chromosome analysis (107) to detect a *de novo* cytogenetically balanced translocation between chromosome 2 and chromosome 9; t(2;9)(q11.2;q34.3). Array CGH was

then used to uncover a submicroscopic 2.7 Mb deletion of a subtelomeric region of 9q34.3, demonstrating that the rearrangement was unbalanced. The results were confirmed using FISH. Effective analysis of dosage-sensitive genomic regions is of high importance to prenatal care.

In a larger study, 300 prenatal samples were analyzed. The most common indications were advanced maternal age and abnormal ultrasound findings. Copy number changes were detected in approximately 58 (19%) of the samples, with only 15 (5%) bearing pathological significance (108). They concluded that aCGH has improved diagnostics for prenatal chromosomal tests.

5.5. Apparently Balanced, Unbalanced Translocations

The use of DNA microarrays has shown that cytogenetically visible, apparently "balanced translocations" can actually be unbalanced and present complex rearrangements, like deletions, inversions, and insertions at or near one or both breakpoints. One study showed that six out of ten patients with abnormal phenotypes but what was thought to be balanced translocations presented previously unrecognized imbalances on the chromosomes involved in the translocation (117). Another study compared phenotypically normal and abnormal individuals that were *balanced* translocation carriers and found that in all affected individuals, there were genomic imbalances at the breakpoints or elsewhere (118).

5.6. Recurrent Translocations

NAHR is mostly thought of as a mechanism that leads to interstitial microdeletions, microduplications, and inversions, but translocations might also be stimulated by genomic architectural features found on different chromosomes. Such seems to be the case for two of the recurrent constitutional translocations t(11;22)(q23;q11), in which palindromic AT rich repeats (PATTR) are responsible for the rearrangement (119–124) and t(4;8)(p16;p23), in which olfactory receptor-gene cluster LCRs mediate the translocation via NAHR (125, 126).

A more recent publication studied three patients with an unbalanced translocation der(4)t(4;11)(p16.2;p15.4). In all three cases, they found that the breakpoints occurred within LCRs that extended over 200 kb in length and had over 94% DNA sequence identity. Further analysis of two other previously reported patients with the same translocation documented that in all (five out of five) recurrent t(4;11) translocations, the rearrangement had occurred by NAHR, mediated by interchromosomal LCRs. Subsequently, using computational methods, they analyzed the genome-wide presence of interchromosomal LCRs greater than 10 kb in length, and a "recurrent translocation map" was constructed. The map predicted the existence of ~400 interchromosomal LCRs (20 kb in length) potentially able to act as HR substrates and mediate recurrent translocations by NAHR. Upon the generation of the "recurrent translocation map," the authors reevaluated the public and clinical

laboratory databases and found six other examples of recurrent translocations which reported breakpoints predicted by the map (127). These results suggest that human genomic architecture plays a role in recurrent translocations.

5.7. Insertional Translocations

When interpreting aCGH results, one must always remember that a comparative analysis for copy number is performed, and that arrays do not provide information regarding position or orientation of the change. The term "gain" is preferred over "duplication" because in some cases the genomic interval showing apparent gain (when compared to control) is not inserted adjacent to the original sequence. In some instances, segments that become duplicated might be inserted into another genomic location, a phenomenon known as insertional translocation. These events are considered rare as they require that at least three breaks occur (as opposed to one or two breaks required for deletions, duplications, or terminal translocations). In this situation, the use of microarrays is insufficient to confirm the location of the duplicated sequence, and FISH is required to identify the location of the additional material (128).

The previously estimated frequency of insertional translocation was about 1:80,000 karyotypes (129, 130). Using aCGH together with FISH analyses, a total of 40/18,000 cases were identified in which insertional translocation had occurred (a frequency of about 1:500) 160 times greater than previously established in the literature using techniques of limited resolution (128).

5.8. Complex Rearrangements

Genomic rearrangements can be complex. Examples include triplications within duplications, noncontiguous duplications, insertions at the breakpoint junction of deletions and duplications, etc. Complex rearrangements are now being found in numerous loci across the genome (reviewed by Zhang et al. (131)).

PMD, an X-linked dysmyelinating disorder of the central nervous system is caused by point mutations or genomic rearrangements involving the *PLP1* gene. Previously, it was thought that the PMD associated with *PLP1* duplication occurred mainly through coupled homologous and nonhomologous recombination mechanisms (66, 132). In 2007, a study (83) used high resolution arrays and breakpoint sequencing to study the rearrangements of 17 patients with *PLP1* duplications; remarkably, 65% of the samples presented complexities, such as interspersed stretches of DNA of normal copy number amid the duplications, triplications within duplications, insertions at the breakpoint junctions, etc.

Deletions in 17p13.3 cause Miller–Dieker syndrome. Individuals with submicroscopic duplication of the same region present an increased risk for macrosomia, mild developmental delay, and pervasive developmental disorder with a characteristic facial dysmorphology. In 2009, a report described seven patients

with microduplications, all of which were nonrecurrent, and three (about 42%) presented complex rearrangements (133).

Duplication of the *MECP2* gene on Xq28 is one of the most common rearrangements identified in males that present developmental delay and the most common subtelomeric duplication (37). Xq28 shows intricate genomic architecture and presents multiple direct and inverted LCRs, causing both polymorphic structural variation and disease in the population (71, 134, 135). Using a 4-Mb tiling-path oligonucleotide array, an analysis of 30 patients with duplications, including the *MECP2* gene (134), reported finding nonrecurrent rearrangements that ranged in size from 250 kb to 2.6 Mb; complex rearrangements were found in 27% of patients (six subjects showed triplications within the duplications and two samples with stretches of nonduplicated segments embedded in the duplicated region). Complex rearrangements in these regions were not described previously as karyotype analysis and/or BAC arrays did not provide the necessary level of human genome resolution. The use of tiling oligonucleotide arrays has enabled reevaluation of many patient samples and suggests that complexity is prevalent in numerous nonrecurrent rearrangements that cause genomic disorders (134).

Using high-density aCGH, two sample sets with rearrangements involving 17p11.2 or 17p12 were analyzed. The first group consisted of 14 nonrecurrent PTLS-associated duplications (17p11.2) that varied in size from 3.5 to 19.6 Mb. Interestingly, complex rearrangements were found in over half of those patients (57%). The second group was made of seven samples with previously identified rearrangements involving the CMT1A/HNPP region (17p12) bearing multiplex ligation-dependent probe amplification patterns inconsistent with rearrangements mediated by NAHR. One sample was shown to have a complex rearrangement and the other six showed exonic deletions. In both cohorts, the data shows that complex rearrangements may be generated by FoSTeS/MMBIR, that these can range in size from a few base pairs to about 20 Mb, and that rearrangements mediated by FoSTeS can originate mitotically (136).

When PTLS-associated nonrecurrent rearrangements were investigated, over 50% were shown to be complex rearrangements. Using aCGH and breakpoint sequence analysis a study analyzing 21 individuals with nonrecurrent *PMP22* CNVs revealed that various mechanisms (NHEJ, *Alu–Alu* mediated recombination FoSTeS/MMBIR) could be responsible for generating the nonrecurrent 17p12 rearrangements associated with neuropathy. Among the 21 patients, three (~14%) presented deletions that involved one or more exons of the *PMP22* gene. Seven other patients presented partial *PMP22* deletions, showing that partial *PMP22* deletion can also result in loss-of-function mutations and haploinsufficiency of the PMP22 protein, causing neuropathy (82).

5.9. Exonic CNVs

Rearrangements have often been considered as deletions, duplications, inversions, or translocations of large blocks of DNA, but it was recently shown (136) that this might not always be the case since rearrangements can involve single exons. In some examples presented in Subheading 5.8, authors report finding small rearrangements that included one or more exons: Zhang et al. (136) reported 86% of samples with rearrangements in 17p12 had exonic CNVs. When analyzing *PMP22* CNVs, 34% showed rearrangements involving one or a few exons (82).

Using 180 K exon-targeted arrays to examine 2,550 samples, Cheung et al. reported 15 cases (0.59%) of intragenic rearrangements involving exons of different genes (*FMA58A*, *PTEN*, *CREBBP*, *DLG3*, among others). The phenotypes of the patients were consistent with the syndrome described for the affected gene. Such findings would have been missed using non-exon-targeted aCGH and other molecular diagnostic methods, reaffirming the importance of using high-resolution techniques to evaluate patients with unexplained mental retardation and congenital anomalies (137, 138).

6. Final Remarks

Throughout this chapter, the methods used to study genomic rearrangements related to human diseases were described (128) (chromosome banding, FISH, microarrays), and their ability to resolve changes in the human genome of increasingly smaller sizes, including exons of only a few hundred base pairs. The use of high resolution human genome analysis in clinical diagnostics, and its limitations (e.g., positional, orientational information) have been enumerated. Interestingly, there is a direct relationship between the technological advancements and the ability to resolve genome changes and the discovery and description of the mechanisms (NAHR, NHEJ, and FoSTeS/MMBIR) that may be causal to such genomic changes, and hence the understanding of our genome. We now know that NAHR typically leads to recurrent rearrangements of the same size in different patients, which is expected considering that this mechanism reflects the architecture of the region and the requirement for HR substrates. On the other hand, nonrecurrent rearrangements can be generated by NHEJ or FoSTeS/MMBIR, the latter responsible for producing complex alterations (e.g., triplications within duplications), which were seldom reported before the use of high-resolution CGH. It has also been suggested that FoSTeS/MMBIR may have a role in the molecular evolutionary process underlying events, such as exon shuffling, gene fusion/fission, and exon accretion, and thus represent an important mechanism for evolving our genome.

References

1. Lupski, J. R. (1998) Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits, *Trends Genet 14*, 417–422.
2. Lupski, J. R. (2009) Genomic disorders ten years on, *Genome Med 1*, 42.
3. Nahon, J. L. (2003) Birth of 'human-specific' genes during primate evolution, *Genetica 118*, 193–208.
4. Bailey, J. A., and Eichler, E. E. (2006) Primate segmental duplications: crucibles of evolution, diversity and disease, *Nat Rev Genet 7*, 552–564.
5. Stankiewicz, P., Shaw, C. J., Withers, M., Inoue, K., and Lupski, J. R. (2004) Serial segmental duplications during primate evolution result in complex human genome architecture, *Genome Res 14*, 2209–2220.
6. Fortna, A., Kim, Y., MacLaren, E., *et al.* (2004) Lineage-specific gene duplication and loss in human and great ape evolution, *PLoS Biol 2*, E207.
7. Dumas, L., Kim, Y. H., Karimpour-Fard, A., *et al.* (2007) Gene copy number variation spanning 60 million years of human and primate evolution, *Genome Res 17*, 1266–1277.
8. Lupski, J. R., de Oca-Luna, R. M., Slaugenhaupt, S., *et al.* (1991) DNA duplication associated with Charcot-Marie-Tooth disease type 1A, *Cell 66*, 219–232.
9. Lupski, J. R., and Stankiewicz, P. (2005) Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes, *PLoS Genet 1*, e49.
10. Iafrate, A. J., Feuk, L., Rivera, M. N., *et al.* (2004) Detection of large-scale variation in the human genome, *Nat Genet 36*, 949–951.
11. Korbel, J. O., Urban, A. E., Affourtit, J. P., *et al.* (2007) Paired-end mapping reveals extensive structural variation in the human genome, *Science 318*, 420–426.
12. Sebat, J., Lakshmi, B., Troge, J., *et al.* (2004) Large-scale copy number polymorphism in the human genome, *Science 305*, 525–528.
13. Redon, R., Ishikawa, S., Fitch, K. R., *et al.* (2006) Global variation in copy number in the human genome, *Nature 444*, 444–454.
14. Wong, K. K., deLeeuw, R. J., Dosanjh, N. S., *et al.* (2007) A comprehensive analysis of common copy-number variations in the human genome, *Am J Hum Genet 80*, 91–104.
15. Khaja, R., Zhang, J., MacDonald, J. R., *et al.* (2006) Genome assembly comparison identifies structural variants in the human genome, *Nat Genet 38*, 1413–1418.
16. Newman, T. L., Tuzun, E., Morrison, V. A., *et al.* (2005) A genome-wide survey of structural variation between human and chimpanzee, *Genome Res 15*, 1344–1356.
17. Fiegler, H., Redon, R., Andrews, D., *et al.* (2006) Accurate and reliable high-throughput detection of copy number variation in the human genome, *Genome Res 16*, 1566–1574.
18. Komura, D., Shen, F., Ishikawa, S., *et al.* (2006) Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays, *Genome Res 16*, 1575–1584.
19. Lupski, J. R. (2007) Structural variation in the human genome, *N Engl J Med 356*, 1169–1171.
20. Tuzun, E., Sharp, A. J., Bailey, J. A., *et al.* (2005) Fine-scale structural variation of the human genome, *Nat Genet 37*, 727–732.
21. Turner, D. J., Miretti, M., Rajan, D., *et al.* (2008) Germline rates of *de novo* meiotic deletions and duplications causing several genomic disorders, *Nat Genet 40*, 90–95.
22. Bruder, C. E., Piotrowski, A., Gijsbers, A. A., *et al.* (2008) Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles, *Am J Hum Genet 82*, 763–771.
23. Liang, Q., Conte, N., Skarnes, W. C., and Bradley, A. (2008) Extensive genomic copy number variation in embryonic stem cells, *Proc Natl Acad Sci USA 105*, 17453–17456.
24. Flores, M., Morales, L., Gonzaga-Jauregui, C., *et al.* (2007) Recurrent DNA inversion rearrangements in the human genome, *Proc Natl Acad Sci USA 104*, 6099–6106.
25. Hastings, P. J., Lupski, J. R., Rosenberg, S. M., and Ira, G. (2009) Mechanisms of change in gene copy number, *Nat Rev Genet 10*, 551–564.
26. Britten, R. J., and Kohne, D. E. (1968) Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms, *Science 161*, 529–540.
27. Lander, E. S., Linton, L. M., Birren, B., *et al.* (2001) Initial sequencing and analysis of the human genome, *Nature 409*, 860–921.
28. Edelmann, L., Pandita, R. K., and Morrow, B. E. (1999) Low-copy repeats mediate the common 3-Mb deletion in patients with velo-cardio-facial syndrome, *Am J Hum Genet 64*, 1076–1086.
29. Bailey, J. A., Yavor, A. M., Massa, H. F., Trask, B. J., and Eichler, E. E. (2001) Segmental

duplications: organization and impact within the current human genome project assembly, *Genome Res 11*, 1005–1017.

30. Stankiewicz, P., and Lupski, J. R. (2002) Genome architecture, rearrangements and genomic disorders, *Trends Genet 18*, 74–82.
31. Bacolla, A., and Wells, R. D. (2004) Non-B DNA conformations, genomic rearrangements, and human disease, *J Biol Chem 279*, 47411–47414.
32. Bacolla, A., Jaworski, A., Larson, J. E., *et al.* (2004) Breakpoints of gross deletions coincide with non-B DNA conformations, *Proc Natl Acad Sci USA 101*, 14162–14167.
33. Bacolla, A., Wojciechowska, M., Kosmider, B., Larson, J. E., and Wells, R. D. (2006) The involvement of non-B DNA structures in gross chromosomal rearrangements, *DNA Repair (Amst) 5*, 1161–1170.
34. Yu, X., and Gabriel, A. (2004) Reciprocal translocations in Saccharomyces cerevisiae formed by nonhomologous end joining, *Genetics 166*, 741–751.
35. Elliott, B., Richardson, C., and Jasin, M. (2005) Chromosomal translocation mechanisms at intronic *Alu* elements in mammalian cells, *Mol Cell 17*, 885–894.
36. Oza, P., Jaspersen, S. L., Miele, A., Dekker, J., and Peterson, C. L. (2009) Mechanisms that regulate localization of a DNA double-strand break to the nuclear periphery, *Genes Dev 23*, 912–927.
37. Shao, L., Shaw, C. A., Lu, X. Y., *et al.* (2008) Identification of chromosome abnormalities in subtelomeric regions by microarray analysis: a study of 5,380 cases, *Am J Med Genet A 146A*, 2242–2251.
38. Yatsenko, S. A., Brundage, E. K., Roney, E. K., Cheung, S. W., Chinault, A. C., and Lupski, J. R. (2009) Molecular mechanisms for subtelomeric rearrangements associated with the 9q34.3 microdeletion syndrome, *Hum Mol Genet 18*, 1924–1936.
39. Zhang, L., Lu, H. H., Chung, W. Y., Yang, J., and Li, W. H. (2005) Patterns of segmental duplication in the human genome, *Mol Biol Evol 22*, 135–141.
40. She, X., Horvath, J. E., Jiang, Z., *et al.* (2004) The structure and evolution of centromeric transition regions within the human genome, *Nature 430*, 857–864.
41. Nguyen, D. Q., Webber, C., and Ponting, C. P. (2006) Bias of selection on human copy-number variants, *PLoS Genet 2*, e20.
42. Payen, C., Koszul, R., Dujon, B., and Fischer, G. (2008) Segmental duplications arise from Pol32-dependent repair of broken forks through two alternative replication-based mechanisms, *PLoS Genet 4*, e1000175.
43. Nobile, C., Toffolatti, L., Rizzi, F., *et al.* (2002) Analysis of 22 deletion breakpoints in dystrophin intron 49, *Hum Genet 110*, 418–421.
44. Visser, R., Shimokawa, O., Harada, N., *et al.* (2005) Identification of a 3.0-kb major recombination hotspot in patients with Sotos syndrome who carry a common 1.9-Mb microdeletion, *Am J Hum Genet 76*, 52–67.
45. de Smith, A. J., Walters, R. G., Coin, L. J., *et al.* (2008) Small deletion variants have stable breakpoints commonly associated with alu elements, *PLoS One 3*, e3104.
46. Liskay, R. M., Letsou, A., and Stachelek, J. L. (1987) Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells, *Genetics 115*, 161–167.
47. Reiter, L. T., Hastings, P. J., Nelis, E., De Jonghe, P., Van Broeckhoven, C., and Lupski, J. R. (1998) Human meiotic recombination products revealed by sequencing a hotspot for homologous strand exchange in multiple HNPP deletion patients, *Am J Hum Genet 62*, 1023–1033.
48. Pentao, L., Wise, C. A., Chinault, A. C., Patel, P. I., and Lupski, J. R. (1992) Charcot-Marie-Tooth type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit, *Nat Genet 2*, 292–300.
49. Lam, K. W., and Jeffreys, A. J. (2006) Processes of copy-number change in human DNA: the dynamics of {alpha}-globin gene deletion, *Proc Natl Acad Sci USA 103*, 8921–8927.
50. Lam, K. W., and Jeffreys, A. J. (2007) Processes of *de novo* duplication of human alpha-globin genes, *Proc Natl Acad Sci USA 104*, 10950–10955.
51. Shaw, C. J., and Lupski, J. R. (2005) Non-recurrent 17p11.2 deletions are generated by homologous and non-homologous mechanisms, *Hum Genet 116*, 1–7.
52. Kidd, J. M., Cooper, G. M., Donahue, W. F., *et al.* (2008) Mapping and sequencing of structural variation from eight human genomes, *Nature 453*, 56–64.
53. Gu, W., Zhang, F., and Lupski, J. R. (2008) Mechanisms for human genomic rearrangements, *Pathogenetics 1*, 4.
54. Barbouti, A., Stankiewicz, P., Nusbaum, C., *et al.* (2004) The breakpoint region of the most common isochromosome, i(17q), in human neoplasia is characterized by a complex

genomic architecture with large, palindromic, low-copy repeats, *Am J Hum Genet 74*, 1–10.

55. Carvalho, C. M., and Lupski, J. R. (2008) Copy number variation at the breakpoint region of isochromosome 17q, *Genome Res 18*, 1724–1732.
56. Mendrzyk, F., Korshunov, A., Toedt, G., *et al.* (2006) Isochromosome breakpoints on 17p in medulloblastoma are flanked by different classes of DNA sequence repeats, *Genes Chromosomes Cancer 45*, 401–410.
57. Kehrer-Sawatzki, H., Schmid, E., Funsterer, C., Kluwe, L., and Mautner, V. F. (2008) Absence of cutaneous neurofibromas in an NF1 patient with an atypical deletion partially overlapping the common 1.4 Mb microdeleted region, *Am J Med Genet A 146A*, 691–699.
58. Zhang, F., Potocki, L., Sampson, J. B., *et al.* (2010) Identification of uncommon recurrent Potocki-Lupski syndrome-associated duplications and the distribution of rearrangement types and mechanisms in PTLS, *Am J Hum Genet 86*, 462–470.
59. Krangel, M. S., Hernandez-Munain, C., Lauzurica, P., McMurry, M., Roberts, J. L., and Zhong, X. P. (1998) Developmental regulation of V(D)J recombination at the TCR alpha/delta locus, *Immunol Rev 165*, 131–147.
60. Burma, S., Chen, B. P., and Chen, D. J. (2006) Role of non-homologous end joining (NHEJ) in maintaining genomic integrity, *DNA Repair (Amst) 5*, 1042–1048.
61. Weterings, E., and van Gent, D. C. (2004) The mechanism of non-homologous end-joining: a synopsis of synapsis, *DNA Repair (Amst) 3*, 1425–1435.
62. Lieber, M. R. (2008) The mechanism of human nonhomologous DNA end joining, *J Biol Chem 283*, 1–5.
63. Daley, J. M., Palmbos, P. L., Wu, D., and Wilson, T. E. (2005) Nonhomologous end joining in yeast, *Annu Rev Genet 39*, 431–451.
64. Haviv-Chesner, A., Kobayashi, Y., Gabriel, A., and Kupiec, M. (2007) Capture of linear fragments at a double-strand break in yeast, *Nucleic Acids Res 35*, 5192–5202.
65. Yu, X., and Gabriel, A. (2003) Ku-dependent and Ku-independent end-joining pathways lead to chromosomal rearrangements during double-strand break repair in *Saccharomyces cerevisiae*, *Genetics 163*, 843–856.
66. Lee, J. A., Inoue, K., Cheung, S. W., Shaw, C. A., Stankiewicz, P., and Lupski, J. R. (2006) Role of genomic architecture in *PLP1* duplication causing Pelizaeus-Merzbacher disease, *Hum Mol Genet 15*, 2250–2265.
67. Michel, B., Ehrlich, S. D., and Uzest, M. (1997) DNA double-strand breaks caused by replication arrest, *EMBO J 16*, 430–438.
68. Arlt, M. F., Mulle, J. G., Schaibley, V. M., *et al.* (2009) Replication stress induces genome-wide copy number changes in human cells that resemble polymorphic and pathogenic variants, *Am J Hum Genet 84*, 339–350.
69. Cabrejo, L., Guyant-Marechal, L., Laquerriere, A., *et al.* (2006) Phenotype associated with APP duplication in five families, *Brain 129*, 2966–2976.
70. Chartier-Harlin, M. C., Kachergus, J., Roumier, C., *et al.* (2004) Alpha-synuclein locus duplication as a cause of familial Parkinson's disease, *Lancet 364*, 1167–1169.
71. del Gaudio, D., Fang, P., Scaglia, F., *et al.* (2006) Increased *MECP2* gene copy number as the result of genomic duplication in neurodevelopmentally delayed males, *Genet Med 8*, 784–792.
72. Farrer, M., Kachergus, J., Forno, L., *et al.* (2004) Comparison of kindreds with parkinsonism and alpha-synuclein genomic multiplications, *Ann Neurol 55*, 174–179.
73. Ibanez, P., Bonnet, A. M., Debarges, B., *et al.* (2004) Causal relation between alpha-synuclein gene duplication and familial Parkinson's disease, *Lancet 364*, 1169–1171.
74. Meins, M., Lehmann, J., Gerresheim, F., *et al.* (2005) Submicroscopic duplication in Xq28 causes increased expression of the MECP2 gene in a boy with severe mental retardation and features of Rett syndrome, *J Med Genet 42*, e12.
75. Nishioka, K., Hayashi, S., Farrer, M. J., *et al.* (2006) Clinical heterogeneity of alpha-synuclein gene duplication in Parkinson's disease, *Ann Neurol 59*, 298–309.
76. Polymeropoulos, M. H., Lavedan, C., Leroy, E., *et al.* (1997) Mutation in the alpha-synuclein gene identified in families with Parkinson's disease, *Science 276*, 2045–2047.
77. Potocki, L., Bi, W., Treadwell-Deering, D., *et al.* (2007) Characterization of Potocki-Lupski syndrome (dup(17)(p11.2p11.2)) and delineation of a dosage-sensitive critical interval that can convey an autism phenotype, *Am J Hum Genet 80*, 633–649.
78. Rovelet-Lecrux, A., Hannequin, D., Raux, G., *et al.* (2006) APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy, *Nat Genet 38*, 24–26.
79. Singleton, A. B., Farrer, M., Johnson, J., *et al.* (2003) Alpha-Synuclein locus triplication causes Parkinson's disease, *Science 302*, 841.

80. Van Esch, H., Bauters, M., Ignatius, J., *et al.* (2005) Duplication of the *MECP2* region is a frequent cause of severe mental retardation and progressive neurological symptoms in males, *Am J Hum Genet 77*, 442–453.
81. Vissers, L. E., Stankiewicz, P., Yatsenko, S. A., *et al.* (2007) Complex chromosome 17p rearrangements associated with low-copy repeats in two patients with congenital anomalies, *Hum Genet 121*, 697–709.
82. Zhang, F., Seeman, P., Pengfei, L., *et al.* (2010) Mechanisms for nonrecurrent genomic rearrangements associated with CMT1A or HNPP: rare CNVs as a cause for missing heritability, *Am J Hum Genet 86*, 892–903.
83. Lee, J. A., Carvalho, C. M., and Lupski, J. R. (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders, *Cell 131*, 1235–1247.
84. Hastings, P. J., Ira, G., and Lupski, J. R. (2009) A microhomology-mediated break-induced replication model for the origin of human copy number variation, *PLoS Genet 5*, e1000327.
85. Hastings, P. J., Bull, H. J., Klump, J. R., and Rosenberg, S. M. (2000) Adaptive amplification: an inducible chromosomal instability mechanism, *Cell 103*, 723–731.
86. Hastings, P. J., Slack, A., Petrosino, J. F., and Rosenberg, S. M. (2004) Adaptive amplification and point mutation are independent mechanisms: evidence for various stress-inducible mutation mechanisms, *PLoS Biol 2*, e399.
87. Slack, A., Thornton, P. C., Magner, D. B., Rosenberg, S. M., and Hastings, P. J. (2006) On the mechanism of gene amplification induced under stress in *Escherichia coli*, *PLoS Genet 2*, e48.
88. Ponder, R. G., Fonville, N. C., and Rosenberg, S. M. (2005) A switch from high-fidelity to error-prone DNA double-strand break repair underlies stress-induced mutation, *Mol Cell 19*, 791–804.
89. Merrihew, R. V., Marburger, K., Pennington, S. L., Roth, D. B., and Wilson, J. H. (1996) High-frequency illegitimate integration of transfected DNA at preintegrated target sites in a mammalian genome, *Mol Cell Biol 16*, 10–18.
90. Borg, I., Delhanty, J. D., and Baraitser, M. (1995) Detection of hemizygosity at the elastin locus by FISH analysis as a diagnostic test in both classical and atypical cases of Williams syndrome, *J Med Genet 32*, 692–696.
91. Yatsenko, S. A., Shaw, C. A., Ou, Z., *et al.* (2009) Microarray-based comparative genomic hybridization using sex-matched reference DNA provides greater sensitivity for detection of sex chromosome imbalances than array-comparative genomic hybridization with sex-mismatched reference DNA, *J Mol Diagn 11*, 226–237.
92. Beaudet, A. L., and Belmont, J. W. (2008) Array-based DNA diagnostics: let the revolution begin, *Annu Rev Med 59*, 113–129.
93. Stankiewicz, P., and Beaudet, A. L. (2007) Use of array CGH in the evaluation of dysmorphology, malformations, developmental delay, and idiopathic mental retardation, *Curr Opin Genet Dev 17*, 182–192.
94. Alkan, C., Kidd, J. M., Marques-Bonet, T., *et al.* (2009) Personalized copy number and segmental duplication maps using next-generation sequencing, *Nat Genet 41*, 1061–1067.
95. Cheung, S. W., Shaw, C. A., Scott, D. A., *et al.* (2007) Microarray-based CGH detects chromosomal mosaicism not revealed by conventional cytogenetics, *Am J Med Genet A 143A*, 1679–1686.
96. Ballif, B. C., Rorem, E. A., Sundin, K., *et al.* (2006) Detection of low-level mosaicism by array CGH in routine diagnostic specimens, *Am J Med Genet A 140*, 2757–2767.
97. Lu, X. Y., Phung, M. T., Shaw, C. A., *et al.* (2008) Genomic imbalances in neonates with birth defects: high detection rates by using chromosomal microarray analysis, *Pediatrics 122*, 1310–1318.
98. Scott, S. A., Cohen, N., Brandt, T., Toruner, G., Desnick, R. J., and Edelmann, L. (2010) Detection of low-level mosaicism and placental mosaicism by oligonucleotide array comparative genomic hybridization, *Genet Med 12*, 85–92.
99. Piotrowski, A., Bruder, C. E., Andersson, R., *et al.* (2008) Somatic mosaicism for copy number variation in differentiated human tissues, *Hum Mutat 29*, 1118–1124.
100. Linardopoulou, E. V., Williams, E. M., Fan, Y., Friedman, C., Young, J. M., and Trask, B. J. (2005) Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication, *Nature 437*, 94–100.
101. Riethman, H., Ambrosini, A., and Paul, S. (2005) Human subtelomere structure and variation, *Chromosome Res 13*, 505–515.
102. Cheung, S. W., Shaw, C. A., Yu, W., *et al.* (2005) Development and validation of a CGH microarray for clinical cytogenetic diagnosis, *Genet Med 7*, 422–432.
103. Ravnan, J. B., Tepperberg, J. H., Papenhausen, P., *et al.* (2006) Subtelomere FISH analysis of 11 688 cases: an evaluation of the frequency and pattern of subtelomere rearrangements in

individuals with developmental disabilities, *J Med Genet 43*, 478–489.

104. Ballif, B. C., Sulpizio, S. G., Lloyd, R. M., *et al.* (2007) The clinical utility of enhanced subtelomeric coverage in array CGH, *Am J Med Genet A 143A*, 1850–1857.
105. Rickman, L., Fiegler, H., Shaw-Smith, C., *et al.* (2006) Prenatal detection of unbalanced chromosomal rearrangements by array CGH, *J Med Genet 43*, 353–361.
106. Bi, W., Breman, A. M., Venable, S. F., *et al.* (2008) Rapid prenatal diagnosis using uncultured amniocytes and oligonucleotide array CGH, *Prenat Diagn 28*, 943–949.
107. Simovich, M. J., Yatsenko, S. A., Kang, S. H., *et al.* (2007) Prenatal diagnosis of a 9q34.3 microdeletion by array-CGH in a fetus with an apparently balanced translocation, *Prenat Diagn 27*, 1112–1117.
108. Van den Veyver, I. B., Patel, A., Shaw, C. A., *et al.* (2009) Clinical use of array comparative genomic hybridization (aCGH) for prenatal diagnosis in 300 cases, *Prenat Diagn 29*, 29–39.
109. Lupski, J. R. (2007) Genomic rearrangements and sporadic disease, *Nat Genet 39*, S43-47.
110. Jacobs, P. A., Browne, C., Gregson, N., Joyce, C., and White, H. (1992) Estimates of the frequency of chromosome abnormalities detectable in unselected newborns using moderate levels of banding, *J Med Genet 29*, 103–108.
111. Shah, V. C., Murthy, D. S., and Murthy, S. K. (1990) Cytogenetic studies in a population suspected to have chromosomal abnormalities, *Indian J Pediatr 57*, 235–243.
112. Kenue, R. K., Raj, A. K., Harris, P. F., and el-Bualy, M. S. (1995) Cytogenetic analysis of children suspected of chromosomal abnormalities, *J Trop Pediatr 41*, 77–80.
113. Goud, M. T., Al-Harassi, S. M., Al-Khalili, S. A., Al-Salmani, K. K., Al-Busaidy, S. M., and Rajab, A. (2005) Incidence of chromosome abnormalities in the Sultanate of Oman, *Saudi Med J 26*, 1951–1957.
114. Kim, S. S., Jung, S. C., Kim, H. J., Moon, H. R., and Lee, J. S. (1999) Chromosome abnormalities in a referred population for suspected chromosomal aberrations: a report of 4117 cases, *J Korean Med Sci 14*, 373–376.
115. Warburton, D. (1991) *De novo* balanced chromosome rearrangements and extra marker chromosomes identified at prenatal diagnosis: clinical significance and distribution of breakpoints, *Am J Hum Genet 49*, 995–1013.
116. Sahoo, T., Cheung, S. W., Ward, P., *et al.* (2006) Prenatal diagnosis of chromosomal abnormalities using array-based comparative genomic hybridization, *Genet Med 8*, 719–727.
117. Gribble, S. M., Prigmore, E., Burford, D. C., *et al.* (2005) The complex nature of constitutional *de novo* apparently balanced translocations in patients presenting with abnormal phenotypes, *J Med Genet 42*, 8–16.
118. Baptista, J., Mercer, C., Prigmore, E., *et al.* (2008) Breakpoint mapping and array CGH in translocations: comparison of a phenotypically normal and an abnormal cohort, *Am J Hum Genet 82*, 927–936.
119. Zackai, E. H., and Emanuel, B. S. (1980) Site-specific reciprocal translocation, t(11;22)(q23;q11), in several unrelated families with 3:1 meiotic disjunction, *Am J Med Genet 7*, 507–521.
120. Shaikh, T. H., Budarf, M. L., Celle, L., Zackai, E. H., and Emanuel, B. S. (1999) Clustered 11q23 and 22q11 breakpoints and 3:1 meiotic malsegregation in multiple unrelated t(11;22) families, *Am J Hum Genet 65*, 1595–1607.
121. Kurahashi, H., and Emanuel, B. S. (2001) Long AT-rich palindromes and the constitutional t(11;22) breakpoint, *Hum Mol Genet 10*, 2605–2617.
122. Edelmann, L., Spiteri, E., Koren, K., *et al.* (2001) AT-rich palindromes mediate the constitutional t(11;22) translocation, *Am J Hum Genet 68*, 1–13.
123. Ashley, T., Gaeth, A. P., Inagaki, H., *et al.* (2006) Meiotic recombination and spatial proximity in the etiology of the recurrent t(11;22), *Am J Hum Genet 79*, 524–538.
124. Kato, T., Inagaki, H., Yamada, K., *et al.* (2006) Genetic variation affects *de novo* translocation frequency, *Science 311*, 971.
125. Giglio, S., Calvari, V., Gregato, G., *et al.* (2002) Heterozygous submicroscopic inversions involving olfactory receptor-gene clusters mediate the recurrent t(4;8)(p16;p23) translocation, *Am J Hum Genet 71*, 276–285.
126. Maas, N. M., Van Vooren, S., Hannes, F., *et al.* (2007) The t(4;8) is mediated by homologous recombination between olfactory receptor gene clusters, but other 4p16 translocations occur at random, *Genet Couns 18*, 357–365.
127. Ou, Z., Stankiewicz, P., Xia, Z., *et al.* (2011) Observation and prediction of recurrent human translocations mediated by NAHR between nonhomologous chromosomes. *Genome Res 21*, 33–46.
128. Kang, S. H., Shaw, C., Ou, Z., *et al.* (2010) Insertional translocation detected using FISH confirmation of array-comparative genomic hybridization (aCGH) results, *Am J Med Genet A 152A*, 1111–1126.

129. Fryns, J. P., Kleczkowska, A., and Kenis, H. (1984) *De novo* complex chromosomal rearrangement (CCR) in a severely mentally retarded boy, *Ann Genet 27*, 62–64.
130. Van Hemel, J. O., and Eussen, H. J. (2000) Interchromosomal insertions. Identification of five cases and a review, *Hum Genet 107*, 415–432.
131. Zhang, F., Carvalho, C. M., and Lupski, J. R. (2009) Complex human chromosomal and genomic rearrangements, *Trends Genet 25*, 298–307.
132. Woodward, K. J., Cundall, M., Sperle, K., *et al.* (2005) Heterogeneous duplications in patients with Pelizaeus-Merzbacher disease suggest a mechanism of coupled homologous and nonhomologous recombination, *Am J Hum Genet 77*, 966–987.
133. Bi, W., Sapir, T., Shchelochkov, O. A., *et al.* (2009) Increased *LIS1* expression affects human and mouse brain development, *Nat Genet 41*, 168–177.
134. Carvalho, C. M., Zhang, F., Liu, P., *et al.* (2009) Complex rearrangements in patients with duplications of *MECP2* can occur by fork stalling and template switching, *Hum Mol Genet 18*, 2188–2203.
135. Small, K., Iber, J., and Warren, S. T. (1997) Emerin deletion reveals a common X-chromosome inversion mediated by inverted repeats, *Nat Genet 16*, 96–99.
136. Zhang, F., Khajavi, M., Connolly, A. M., Towne, C. F., Batish, S. D., and Lupski, J. R. (2009) The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans, *Nat Genet 41*, 849–853.
137. Boone, P. M., Bacino, C. A., Shaw, C. A., *et al.* Detection of clinically relevant exonic copy-number changes by array CGH, *Hum Mutat 31*, 1326–1342.
138. Conrad, D. F., Pinto, D., Redon, R., *et al.* (2009) Origins and functional impact of copy number variation in the human genome, *Nature 464*, 704–712.