Genome architecture, rearrangements and genomic disorders

Pawel Stankiewicz and James R. Lupski

An increasing number of human diseases are recognized to result from recurrent DNA rearrangements involving unstable genomic regions. These are termed genomic disorders, in which the clinical phenotype is a consequence of abnormal dosage of gene(s) located within the rearranged genomic fragments. Both inter- and intrachromosomal rearrangements are facilitated by the presence of region-specific low-copy repeats (LCRs) and result from nonallelic homologous recombination (NAHR) between paralogous genomic segments. LCRs usually span ~10–400 kb of genomic DNA, share ≥97% sequence identity, and provide the substrates for homologous recombination, thus predisposing the region to rearrangements. Moreover, it has been suggested that higher order genomic architecture involving LCRs plays a significant role in karyotypic evolution accompanying primate speciation.

The concept of ‘genomic disorders’ as proposed and defined in 1998 [1] refers to conditions that result from DNA rearrangements due to regional genomic architecture. These rearrangements lead to the loss or gain of a dosage-sensitive gene (or genes) or to disruption of a gene [1]. Such genomic abnormalities have been shown to result mostly from nonallelic homologous recombination (NAHR) between region-specific low-copy repeats (LCRs) [1]. Since that review of the literature, several more diseases have been attributed to genomic rearrangements [2–5], supporting further the concept of genomic disorders and documenting clearly that NAHR is a major mechanism for human disease. An increasing body of data on the human genome primary sequence and higher order architecture has enabled further characterization of known LCRs and the identification of novel LCRs. Several studies indicate that LCRs have a complex structure, have arisen during primate speciation, and appear to be evolving still, thus revealing the plasticity of our genome.

In contrast to conventional monogenic disorders that are due to specific mutations within a gene and reflect errors of DNA replication and/or repair, genomic disorders are recombination-based conditions. Because LCRs can provide large regions of sequence similarity/identity as substrates for homologous recombination, NAHR might not be recognized readily as aberrant by the cellular recombination machinery. This has been proposed as a possible explanation for the high frequency of new mutations in genomic disorders [3].

Genome architecture
Recent estimates [5–9] suggest that as much as 5–10% of the human genome might be duplicated. Genome-wide fluorescence in situ hybridization (FISH) studies using bacterial artificial chromosome (BAC) clones showed that ~5.4% of the human genome is duplicated [7]. Electronic analysis of available sequencing data revealed that ~10% of genomic segments of ≥1 kb and >98% sequence identity are present in at least two copies [9], and further analysis of an updated draft genome suggests that ~5–7% is duplicated [10]. More recently, a computer-based BLAST analysis of electronic data released from the draft sequence of the Human Genome Project revealed potential LCRs that await experimental confirmation [6,9,11]. Bioinformatic analysis of these ‘virtual LCRs’ and their contribution to genome architecture might identify additional regions that are prone to rearrangements that cause genomic disorders. These estimates of the number of duplicated segments will probably change as more finished sequence of the human genome is obtained, because repeat-rich regions present challenges for hybridization-based physical mapping and computational assemblies, and there are limitations in the ability of the draft genome to provide positional information [12]. Erroneous assembly of closely related sequences from nonoverlapping clones, and misassignment of genomic clones, are likely to underestimate the frequency of such genome architectural features. This is true especially for LCRs of 150–200 kb in length, which are approximately the same size as the BAC clones that are the ‘currency’ of the human genome sequencing effort [1]. Misassembly is particularly problematic among genomic segments with the highest sequence similarity [6,8].

Region-specific LCRs, also called segmental duplications or duplons (although this latter term can be ambiguous when more than two copies exist), consist usually of DNA blocks of ~10–400 kb with ≥97% identity that are thought to have arisen by duplication of genomic segments resulting in PARALOGOUS (see Glossary) regions. LCRs can contain genes, gene fragments, pseudogenes, endogenous retroviral sequences or other paralogous fragments. Alternatively, they can contain a series of genes, in which case they represent repeat gene dusters. LCRs are distinguished from highly repetitive sequences in the human genome that were identified on the basis of reassociation kinetics because of the high degree of sequence similarity and large...
Reference

**Paralogous:** Describes two distinct nonallelic genomic segments, with highly similar primary DNA sequence, that are derived from a duplication event.

**Gene conversion:** The alteration of one strand of a heteroduplex DNA to make it complementary with the other strand at any position(s) where there are mispaired bases.

**Homogenization:** The process whereby repeated sequences become more like each other, usually through gene conversion.

Reference


The vast majority of genomic disorders result from submicroscopic chromosome rearrangements, thus limiting their detection by routine cytogenetic analysis, including high-resolution chromosome banding techniques. However, application of FISH analysis to either metaphase or interphase chromosomes has allowed more sensitive detection of the products of aberrant recombination. Pulsed-field gel electrophoresis (PFGE) has also been applied successfully to identify genome rearrangements that were too small to visualize by routine cytological techniques (<3 Mb) yet too large to be resolved by conventional agarose gel electrophoresis (>20 kb) [3]. Disease-specific junction fragments identified by PFGE indicate that the rearrangements arise from a precise recombination mechanism because the identical sized fragment is generated in multiple independent events. Furthermore, these junction fragments provide useful diagnostic tools.

Methods

The vast majority of genomic disorders result from submicroscopic chromosome rearrangements, thus limiting their detection by routine cytogenetic analysis, including high-resolution chromosome banding techniques. However, application of FISH analysis to either metaphase or interphase chromosomes has allowed more sensitive detection of the products of aberrant recombination. Pulsed-field gel electrophoresis (PFGE) has also been applied successfully to identify genome rearrangements that were too small to visualize by routine cytological techniques (<3 Mb) yet too large to be resolved by conventional agarose gel electrophoresis (>20 kb) [3]. Disease-specific junction fragments identified by PFGE indicate that the rearrangements arise from a precise recombination mechanism because the identical sized fragment is generated in multiple independent events. Furthermore, these junction fragments provide useful diagnostic tools.

Rearrangement mechanisms

Substrates for homologous recombination identified to date consist of significant lengths of sequence homology (LCRs), usually ~10–400 kb in size. Analysis of DNA rearrangements associated with genomic disorders reveals that the chromosome/chromatid misalignment that enables nonallelic LCRs to pair as substrates for homologous recombination (NAHR), appears to depend on genome architectural features. These features include repeat size, degree of homology, distance between LCRs, and orientation with respect to each other. It might be that the LCRs’ length, rather than their degree of homology, influences the initial interaction of sequence substrates for recombination. However, it is unclear why in some genomic disorders the majority of strand exchanges or crossovers are restricted to a specific region (positional recombination hotspots) within the repeat [15–18]. The size of chromosome rearrangements also was observed to correlate with the size of repeat: the longer the LCRs, the larger the size of the rearranged genomic fragment [1].

Strand exchange during somatic homologous recombination in mammalian cells appears to require a minimum of ~200–300 bp of uninterrupted homology, which is thought to reflect a minimal efficient processing segment (MEPS) required by the cellular recombination machinery [19]. Whether meiotic homologous recombination requires larger or smaller stretches of identity is unclear. However, direct sequencing of meiotic recombination products suggests that 300–500 bp of identity is required [16]. Recombination products that reveal evidence for gene conversion are consistent with the occurrence of double-strand break events [16,20,21]. Many copies of mariner transposon-like elements and other potential cis-acting recombinagenic sequences are found throughout the human genome and might act as the initiation sites for double-strand DNA breaks [22]. It has also been suggested that the open chromatin conformation associated with active transcription might facilitate homologous recombination [23]. In some cases, double equal crossover or noncrossover conversion events within an LCR could lead to its homogenization. These processes increase the identity between paralogous sequences, enhancing their ability to serve as premutation NAHR substrates [24,25]. In other cases, such recombination might diversify (rather than homogenize) sequences, resulting in either the creation of pseudogenes or disease-associated mutations from gene conversion with pseudogenes [26,27].

It has been proposed that unequal crossovers between directly oriented LCRs on homologous chromosomes can produce two reciprocal products: a direct (tandem) duplication and a deletion [28,29] (Fig. 1a). Unequal crossing-over between nonallelic, directly repeated, homologous segments located on homologous chromosomes (interchromosomal) or sister chromatids (intrachromosomal) has been shown in patients with chromosome duplications involving the Charcot–Marie–Tooth disease type 1A (CMT1A) region in 17p12 [15,28,30–32], dup(17)(p11.2p11.2) [33], and dup(22)(q11.2q11.2) [34] (Fig. 1a,d,f). Mispairing between inverted repeats on homologous chromosomes results in an inversion (Fig. 1b) when crossover involves both repeats [1–4,25,35,36].

Some LCRs have complex structures with sequences among the LCRs oriented in a direct
manner whereas others are inverted. Thus, when considering a specific set of LCRs to be utilized as NAHR substrates, the finished sequence needs to be elucidated to determine the orientation of recombination substrates and to predict the consequences of recombination events. In addition, the number of subsequent crossovers can determine the derivative chromosome products.

Chromosome microdeletions of intrachromosomal origin have been hypothesized to result from one of three LCR/NAHR-based mechanisms. Similar to interchromosomal rearrangement, unequal crossing-over between direct LCRs on sister chromatids is predicted to result in deletion/duplication derivative chromosomes [3,37] (Fig. 1d,f). The intrachromatid foldback loop mediated by directly oriented repeats, followed by a crossing-over event, leads to the loop excision and deletion (Fig. 1g) [34,37]. Intrachromatid recombination events between LCRs can lead to an inversion when the LCRs are in an inverted orientation and NAHR occurs within a single chromatid – for example inv dup(8p) [1,4,25,35,36] (Fig. 1h). It has been proposed that duplicated modules in an inverse orientation can form a ‘stem-loop’ intermediate. Intrachromatid recombination between the duplicated modules that form the ‘stem’ would result in deletion; however, this would require nonhomologous recombination because conventional homologous recombination models would predict inversion [38–40]. Finally, the complex structure of some LCRs can lead to both inversion (Fig. 1i) or deletion and duplication (Fig. 1f), as well as rearrangement of repeat structures within the LCR, depending upon which portion of the LCR is utilized as the recombination substrate [3,25,35,36]. Other chromosomal rearrangements might not be dependent on substrate homology for recombination. The recurrent reciprocal constitutional translocation t(11;22) in humans has been shown to occur via nonhomologous recombination using palindromic AT-rich sequences on 11q23 and 22q11 (within one of the LCR22s), forming a hairpin-like structure [41,42].
Interestingly, characterization of breakpoints at the nucleotide sequence level, in CMT1A/HNPP (hereditary neuropathy with liability to pressure palsies) and neurofibromatosis type 1 (NF1), has demonstrated regions of strand exchange at recombination hotspots within 557 bp and 2 kb, respectively [16–18,21]. With the exception of CMT1A, spinal muscular atrophy (SMA) and NF1, no parent-of-origin bias has been identified among patients with other genomic disorders. About 87% of CMT1A duplication and ~85% of SMA deletion events arise during spermatogenesis [43–45], whereas ~80% of NF1 deletions are of maternal origin [46,47].

The identification of somatic mosaicism for some LCR-based deletions [48–51] as well as reversion of duplication [51,52] indicates that NAHR can also occur in mitosis, suggesting that NAHR might play a significant role in loss of heterozygosity in tumorigenesis [53].

Genomic disorders can manifest as mendelian traits or chromosomal disorders

Genome rearrangements often do not represent random events, but rather reflect higher order genome architectural features that facilitate NAHR. Depending on the size of the genomic segment involved (i.e. the distance between LCRs), and the number of potential dosage-sensitive genes mapping within the rearranged segment, deletions and/or duplications can result in any of the following: a mendelian disease, a contiguous gene syndrome or a chromosomal disorder (Fig. 2). Inversion rearrangements do not alter gene copy number but can disrupt a gene at the junction of the inverted genomic segment. Below we delineate specific examples of representative genomic disorders, emphasizing novel findings and genomic concepts with specific details given in Tables 1–3.

Mendelian diseases and other monogenic traits

Recently, many classical mendelian disorders have been demonstrated to result from genome rearrangements. Table 1 lists the diseases and genome architectural features associated with these DNA rearrangements. Mendelian genomic disorders can segregate as autosomal recessive, autosomal dominant, X-linked or even Y-linked traits.

Autosomal recessive

Familial juvenile nephronophthisis 1 is the most frequent inherited cause of chronic renal failure in children. It is an autosomal recessive trait caused by mutation or deletion of NPHP1 in 2q13, and is most commonly associated with a ~300 kb homozygous deletion on chromosome 2q13 (~80% of cases). The deletion is flanked by large (~330 kb) repeats of inverted orientation [54]. Within and adjacent to the inverted repeat block, a direct repeat of ~45 kb acts as the substrate for NAHR causing the deletion associated with disease [25] (Fig. 3a). A nonpathogenic inversion rearrangement involving the inverted repeats and resulting from NAHR has been identified in 1.3% and in 21% of controls in the homozygous and heterozygous states, respectively [25]. These findings illustrate the high frequency of rearrangements that occur in this genomic region.

Autosomal dominant

In 5–20% of patients with neurofibromatosis type 1 (NF1), interstitial microdeletions of 17q11.2 involving the NF1 gene have been identified. The breakpoints of the common ~1.5 Mbp deletions in the majority of patients cluster at two directly oriented ~85 kb LCR sequences, called proximal and middle NF1-REPs. A third copy, distal NF1-REP, of unknown orientation maps to 17q24 [18]. Polymorphic marker analysis reveals that the microdeletions result predominantly from unequal crossing-over in maternal meiosis I [55] (Fig. 1a).

X-linked

The overwhelming majority (80–85%) of incontinentia pigmenti cases are caused by a recurrent deletion within NEMO on Xq28. The deletions are mediated by directly oriented 870 bp sequence repeats located within intron 3 and a portion of exon 10 [56]. Two-thirds of new mutations originate from the fathers, suggesting a high frequency of intrachromosomal interchanges (Fig. 1d,g).

Y-linked

Two directly oriented ~10 kb copies, HERV15yq1 and HERV15yq2, of the human endogenous retrovirus 15 (HERV15) have been shown to flank the ~800 kb azoospermia (AZF) a microdeletion within Yq11.2 associated with male infertility (Fig. 1d,g). Interestingly, these LCRs contain two hotspots of 1278 bp and 1690 bp within which intrachromosomal recombination results in the majority of AZF a microdeletions. In addition, a
double crossover event between HERV causes the common 12f2 deletion polymorphism [24,57,58]. Recently, the 
~3.5 Mb AZF c region of the Y chromosome has been sequenced. It features complex genome architecture including massive palindromes. The breakpoints of uniform recurrent deletions in infertile men occur within 229 kb direct repeats [59].

Contiguous gene syndromes

Contiguous gene syndromes result from DNA rearrangements (deletion/duplication) that encompass several adjacent genes on a segment of the genome [60,61]. The genome segments involved in several of these conditions have been shown to be flanked by LCRs and the rearrangements mediated by NAHR.

Williams–Beuren syndrome, LCR7

In over 90% of patients with Williams–Beuren syndrome (WBS), the common ~1.6 Mb deletion involving 7q11.23 is mediated by flanking LCR sequences. Three complex repeat gene structures >320 kb in size – referred to as cen, mid and tel, and of ~98% overall homology – are composed of several differentially oriented subunits [62,63] (Fig. 3b). The common deletion results from recombination between directly oriented blocks within the cen and mid LCR7s. Interestingly, inversion of the same segment has been found as a polymorphic variant in parents of WBS patients and in some atypical patients [64].

LCR15

Both the common ~4 Mb deletion and duplication of Prader–Willi syndrome (PWS)/Angelman syndrome (AS) chromosome region have been found to result from unequal exchange between complex LCR15s of >500 kb [39,65–67] (Fig. 3c). More than eight and at most 10–12 copies of complex repeat fragments (50–200 kb) are scattered within 15q11q13; two additional clusters occur within 16p11.2, and there is a further one copy in 15q24. Four large clusters of complex repeats have been termed BP1 to BP4. In ~95% of cases the distal breakpoints of common PWS/AS deletions and 15q11q13 duplications (triplications) map to repeat block LCR15-BP3 consisting of two inverted copies of LCR15-BP3a and LCR15-BP3b. In some unusual larger deletions, the distal breakpoint maps to the repeat block LCR15-BP4. The proximal breakpoints have been found to cluster within LCR15-BP2 (~60%) and within LCR15-BP1 (~40%) [39,66,68,69]. Interestingly, LCR15-BP4 has been shown to be involved also in 15q11q13 triplications [70]. A double U-type exchange event between LCRs15 on both homologous chromosomes and sister

<table>
<thead>
<tr>
<th>Disorders</th>
<th>OMIM*</th>
<th>Inheritance</th>
<th>Chromosome</th>
<th>Gene(s)</th>
<th>Rearrangement</th>
<th>Recombination substrates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bartter syndrome type III</td>
<td>601 678</td>
<td>AD</td>
<td>1p36</td>
<td>CLCNKA/B</td>
<td>del</td>
<td>11</td>
</tr>
<tr>
<td>Gaucher disease</td>
<td>230 800</td>
<td>AR</td>
<td>1q21</td>
<td>GBA</td>
<td>del</td>
<td>16</td>
</tr>
<tr>
<td>Familial juvenile nephronophthisis</td>
<td>256 100</td>
<td>AR</td>
<td>2q13</td>
<td>NPHP1</td>
<td>del</td>
<td>290</td>
</tr>
<tr>
<td>Fascioscapulohumeral muscular dystrophy</td>
<td>158 900</td>
<td>AD</td>
<td>4q35</td>
<td>FRRG1?</td>
<td>del</td>
<td>25–222</td>
</tr>
<tr>
<td>Spinal muscular atrophy</td>
<td>253 300</td>
<td>AR</td>
<td>5q13.2</td>
<td>SMN</td>
<td>inv/dup</td>
<td>500</td>
</tr>
<tr>
<td>Congenital adrenal hyperplasia III/21 hydroxylase deficiency</td>
<td>201 930</td>
<td>AR</td>
<td>6p21.3</td>
<td>CYP21</td>
<td>del</td>
<td>30</td>
</tr>
<tr>
<td>Glucocorticoid-remediable aldosteronism (GRA)</td>
<td>103 900</td>
<td>AD</td>
<td>8q21</td>
<td>CYP11B1/2</td>
<td>dup</td>
<td>45</td>
</tr>
<tr>
<td>β-thalassemia</td>
<td>141 900</td>
<td>AR</td>
<td>11p15.5</td>
<td>β-globin</td>
<td>del</td>
<td>4, (77)</td>
</tr>
<tr>
<td>ξ-thalassemia</td>
<td>141 800</td>
<td>AR</td>
<td>16p13.3</td>
<td>α-globin</td>
<td>del</td>
<td>3.7 or 4.2</td>
</tr>
<tr>
<td>Polycystic kidney disease 1</td>
<td>601 313</td>
<td>AD</td>
<td>16p13.3</td>
<td>PKD1</td>
<td>del</td>
<td>50</td>
</tr>
<tr>
<td>Charcot–Marie–Tooth (CMT1A)</td>
<td>118 220</td>
<td>AD</td>
<td>17p12</td>
<td>PMP22</td>
<td>dup</td>
<td>1400</td>
</tr>
<tr>
<td>Hereditary neuropathy with liability to pressure palsies (HNPP)</td>
<td>162 500</td>
<td>AD</td>
<td>17p12</td>
<td>PMP22</td>
<td>del</td>
<td>1400</td>
</tr>
<tr>
<td>Neurofibromatosis type 1 (NF1)</td>
<td>162 200</td>
<td>AD</td>
<td>17q11.2</td>
<td>NF1</td>
<td>del</td>
<td>1500</td>
</tr>
<tr>
<td>Pituitary dwarfism</td>
<td>262 400</td>
<td>AR</td>
<td>17q23.3</td>
<td>GHI</td>
<td>del</td>
<td>6.7</td>
</tr>
<tr>
<td>CYP2D6 pharmacogenetic trait</td>
<td>124 030</td>
<td>AR</td>
<td>22q13.1</td>
<td>CYP2D6</td>
<td>del/dup</td>
<td>9.3</td>
</tr>
<tr>
<td>Ichthyosis</td>
<td>308 100</td>
<td>XL</td>
<td>Xp22.22</td>
<td>STS</td>
<td>del</td>
<td>1900</td>
</tr>
<tr>
<td>Red–green color blindness</td>
<td>303 800</td>
<td>XL</td>
<td>Xq28</td>
<td>RCP and GCP</td>
<td>del</td>
<td>0</td>
</tr>
<tr>
<td>Incontinentia pigmenti</td>
<td>308 300</td>
<td>XL</td>
<td>Xq28</td>
<td>NEMO</td>
<td>del</td>
<td>10</td>
</tr>
<tr>
<td>Emery–Dreifuss muscular dystrophy</td>
<td>301 300</td>
<td>XL</td>
<td>Xq28</td>
<td>FS</td>
<td>inv</td>
<td>300–500</td>
</tr>
<tr>
<td>Hunter syndrome</td>
<td>309 900</td>
<td>XL</td>
<td>Xq28</td>
<td>Emerin and FLN1</td>
<td>del/dup/inv</td>
<td>11.3</td>
</tr>
</tbody>
</table>

*Abbreviations: AD, autosomal dominant; AR, autosomal recessive; XL, X chromosome linked; del, deletion; dup, duplication; inv, inversion; I, inverted; D, direct; G, gene; ψ, pseudogene; S, segment of genome.


*Molecular evidence demonstrates these conditions result from reciprocal duplication or deletion.
chromatids leading to triplications of proximal 15q has been suggested [69].

LCRs on chromosome 17
Three different LCRs have been identified in chromosome 17: CMT1A-REPs (17p12), Smith–Magenis syndrome-REPs (SMS-REPs) (17p11.2) and NF1-REPs (17q11.2).

Over 90% of patients with SMS harbor a deletion of a ~4 Mb genomic region within chromosome 17p11.2, whereas the remainder of patients carry both smaller and larger deletions. The same genomic segment has been shown to be responsible for a new duplication syndrome, dup(17)(p11.2p11.2) [33]. Physical mapping studies demonstrate the presence of three large ~250 kb region-specific LCR gene clusters termed proximal, middle and distal SMS-REPs [33,71] (Fig. 3d). Preliminary sequence analysis of all three SMS-REPs revealed that the proximal SMS-REP spans ~260 kb and is in the same orientation as the distal REP, which is shorter (~190 kb) and devoid of some short repeat fragments. The middle SMS-REP is inverted with respect to the distal SMS-REP (S-S. Park et al., unpublished). This architecture could explain why the common deletions occur between proximal and distal SMS-REPs. Interestingly, five smaller deletions have been identified, in which the breakpoints are within the inverted middle SMS-REP (P. Stankiewicz and J. R. Lupski, unpublished). In these, the distal breakpoint maps within or adjacent to the distal SMS-REP. It remains to be seen whether a portion of the middle SMS-REP is in direct orientation with respect to the distal SMS-REP (Fig. 3d). SMS-REP-like sequences have been identified on 17p13.1, 17p12, 17q11.2, 17q12, 17q21.2 and 17q23.2, and some appear to be associated with NF1-REP (S-S. Park et al., unpublished).

DiGeorge syndrome, LCR22
Unequal meiotic exchange between LCRs is also a frequent mechanism leading to both common ~3 Mb and smaller ~1.5 Mb deletions within 22q11.2 found in DiGeorge syndrome/velocardiofacial syndrome (DGS/VCFS) [34,40,72]. At least eight LCRs are found on chromosome 22; three are associated with the typically deleted region. These LCRs harbor ~225–400 kb of ~97–98% sequence identity, have complex internal organization and have been termed LCR22-A, LCR22-B, LCR22-C and LCR22-D (ordered centromeric to telomeric) [34,40] (Fig. 3d). LCR22-A and LCR22-D are responsible for the majority (~90%) of the ~3 Mb del(22)(q11.2q11.2) associated with DGS/VCFS and also for the origin of a chromosome duplication, dup(22)(q11.2q11.2) [34,40]. LCR22-A and LCR22-B appear to be utilized as substrates for a smaller unusual-sized (~1.5 Mb)
In almost all cases examined, the breakpoints within LCR15-BP3, LCR15-BP4 and associated with mental retardation, have the distal or Angelman syndrome chromosome region and inv dup(15) containing the Prader–Willi syndrome LCR15-BP1 and LCR15-BP2 (Fig. 3c), whereas larger phenotypically normal individuals involve breakpoints of smaller inv dup(15) found in they cluster within LCRs on chromosome 15. The analysis of inv dup(15) breakpoints reveals that inv dup(15) and inv dup(22), respectively. The bisatellited, pseudodicentric marker chromosomes – are responsible for the origin of two supernumerary LCRs on chromosomes 15 and 22 (LCR15 and LCR22)

Fig. 3. Complex structure of selected low-copy repeats (LCRs). Horizontal lines represent specific genomic regions with the centromere toward the left and telomere to the right. At the right are listed abbreviations for the disease manifested through common deletions of the regions. The colored regions refer to LCRs with the orientation given by the arrowhead. Note complex structure of LCRs consisting of both direct and inverted repeats. (a) LCRs in chromosome 2q13 responsible for rearrangements associated with familial juvenile nephronophthisis 1 (NPHP1). (b) LCRs7 flanking the Williams–Beuren syndrome (WBS) chromosome region 7q11.23. (c) LCRs15 within the Prader–Willi syndrome/Angelman syndrome (PWS/AS) chromosome region 15q11.2. (d) Smith–Magenis syndrome (SMS) repeats within 17p11.2. (e) LCR22 within the DiGeorge syndrome (DGS) chromosome 22q11.2.

deletion found in a few patients. Interestingly, LCR22-B is responsible for der(22)t(11;22) syndrome [41,42], reciprocal t(17;22)(q11;q11) in a family with NF1 [73], and other translocations involving chromosome 22 [74,75].

Marker chromosomes
LCRs on chromosomes 15 and 22 (LCR15 and LCR22) are responsible for the origin of two supernumerary bisatellded, pseudodicentric marker chromosomes – inv dup(15) and inv dup(22), respectively. The analysis of inv dup(15) breakpoints reveals that they cluster within LCRs on chromosome 15. The breakpoints of smaller inv dup(15) found in phenotypically normal individuals involve LCR15-BP1 and LCR15-BP2 (Fig. 3c), whereas larger inv dup(15) containing the Prader–Willi syndrome or Angelman syndrome chromosome region and associated with mental retardation, have the distal breakpoints within LCR15-BP3, LCR15-BP4 and other loci [69,76]. In almost all cases examined, the de novo inv dup(15) occurred during maternal meiosis. The absence of paternally inherited inv dup(15) marker chromosomes might represent an ascertainment bias related to the severity of the phenotype. Alternatively, this might result from parent-of-origin differences of the imprinting status of triplicated genes within the PWS/AS chromosome region and/or it might be related to the prevalence of nondisjunction events in oogenesis. Similar molecular characterization of chromosome breakpoints of inv dup(22) in cat eye syndrome also showed that LCR22-A and LCR22-D (Fig. 3e) are responsible for the marker chromosome associated with cat eye syndrome [34,77]. Based on these data, the majority of inv dup(15) and inv dup(22) chromosomes are thought to occur through a U-type of exchange (allelic breakage and fusion) between homologs involving LCRs on chromosomes 15 and 22, respectively. However, LCRs-based NAHR mechanisms have also been proposed [66,76] (Fig. 1c,e).

Isochromosomes
Until recently, isochromosomes of the long arm of chromosome X (i(Xq)) were thought to be monocentric and result from centromeric misdivision. However, breakpoint analysis reveals that the vast majority of i(Xq)s are in fact dicentric with breakpoints mapping within several duplicated loci in Xp11.21. A U-type breakage/reunion mechanism between sister chromatids or homologous X chromosomes has been proposed for their origin [78]. Alternatively, similar to inv dup(15) and inv dup(22) described above, NAHR between inverted LCRs located in close proximity could be the mechanism responsible for the origin of i(Xq).

Other chromosome rearrangements
Olfactory receptors (ORs) are likely to constitute one of the largest gene superfamilies in the vertebrate genome, comprising ~1% of the DNA length of the human genome [79,80]. Recently, three recurrent, maternally inherited, interrelated genomic rearrangements – inv dup(8p), der(8)(pter8p23.1::p23.2qter) and del(8)(p23.1p23.2) – associated with distinct phenotypes were shown to be mediated by two large OR-LCRs on chromosome 8p. The proximal copy is located on 8p23.1 and the distal copy on 8p23.2. These LCRs span ~400 kb of DNA [36]. A submicroscopic inversion polymorphism was identified in the heterozygous state in 26% of a population of European descent [36]. This inversion heterozygosity was proposed to cause susceptibility to unequal recombination, leading to the origin of three aberrations involving 8p.

Chromosome rearrangements and behavioral traits
Many patients with contiguous gene syndromes and trisomy/tetrasomy of proximal 15q present with
specific behavioral or psychological traits. It has been proposed that some genomic rearrangements can lead to behavioral traits [61]. Indeed, this hypothesis has been supported further by the recent findings that LCR15s-mediated dup(15)(q24q26) appears to be a susceptibility factor for panic and phobic disorders [51]. The observed apparent nonmendelian inheritance has been proposed to be related to the incomplete penetrance resulting from complex interactions between identified mosaicism (mitotic origin of crossovers, reversions, conversions), the variety of distal dup(15q) sizes, and the influence of other genes and epigenetic factors [51].

Conclusions and future prospects

LCR/NAHR-based DNA rearrangements are common, reflect human genome architecture, and can frequently result in genomic disorders. Interestingly, the disorders resulting from genome rearrangements occur with equal frequencies (~10⁻⁴) in distinct world populations. These genomic disorders are due to human genome structural features, in contrast to diseases associated with population-specific alleles. The LCR/NAHR mechanism has been demonstrated to cause mendelian traits, contiguous gene syndromes and whole-arm chromosome aberrations, depending on the size of the genomic segment involved (Fig. 2), indicating this mechanism might play a prominent role in human genetic disease. The evolution of the mammalian genome during primate speciation could have led to the unique genome architecture predisposing some chromosomes to DNA rearrangements that result in genomic disorders [3,81]. To date, many of the DNA rearrangements have been deletions; however, the reciprocal recombination model – CMT1A/HNPP; SMS/dup(17)(p11.2p11.2) – suggests potential duplications from each of these deletion rearrangements. It is very likely that many other submicroscopic genomic rearrangements as well as population-specific polymorphisms mediated by LCRs/NAHRs will be detected in the future, particularly when genome-wide screening methods like high-resolution CGH (comparative genomic hybridization) and human genome BAC-microarrays are applied to genome analysis of patients more commonly [82].

Based on our understanding of genome architecture and susceptibility to DNA rearrangements, we propose that the systematic analysis of the finished human genome sequence data and the identification of genome-wide LCRs will allow the prediction of rearrangement-prone genomic regions. This in turn will enable the development of novel high-resolution genome screening tools for the identification of previously unrecognized genome rearrangements. One such device might be the design of a BAC-microarray with clones from regions flanked by LCRs that can be used for the detection of genomic disorders.

References

7 Cheung, V. G. et al. (2001) Integration of cytogenetic landmarks into the draft sequence of the human genome. Nature 409, 953–958
9 Bailey, J. A. et al. (2001) Segmental duplications: organization and impact within the current human genome project assembly. Genome Res. 11, 1005–1017

30 Inoue, K. et al. (2001) The 1.4-Mb CMT1A duplication/HNPP deletion genomic region reveals unique genome architectural features and provides insights into the recent evolution of new genes. Genome Res. 11, 1018–1033


42 Kamp, C. et al. (2000) Two long homologous retroviral sequence blocks in proximal Yq11 cause AZF microdeletions as a result of intrachromosomal recombination events. Hum. Mol. Genet. 9, 2563–2572


48 Valera, M.C. et al. (2000) Fine-scale comparative mapping of the human 7q11.23 region and the orthologous region on mouse chromosome 5G: the low-copy repeats that flank the Williams–Beuren syndrome deletion are at breakpoint sites of an evolutionary inversion(s). Genomics 69, 1–13


50 Carrozzi, R. et al. (1997) Inter- and intrachromosomal rearrangements are both involved in the origin of the 15q11–q13 deletions in Prader–Willi syndrome. Am. J. Hum. Genet. 61, 228–231


http://tig.trends.com