Human Molecular Genetics, 2011, Vol. 20, No. 10 **1925–1936**
doi:10.1093/hmg/ddr074
Advance Access published on February 24, 2011

FoSTeS, MMBIR and NAHR at the human proximal Xp region and the mechanisms of human Xq isochromosome formation

George Koumbaris[1], Hariklia Hatzisevastou-Loukidou[2], Angelos Alexandrou[1], Marios Ioannides[1], Christodoulos Christodoulou[1], Tomas Fitzgerald[3], Diana Rajan[3], Stephen Clayton[3], Sophia Kitsiou-Tzeli[4], Joris R. Vermeesch[5], Nicos Skordis[6], Pavlos Antoniou[1], Ants Kurg[7], Ioannis Georgiou[8], Nigel P. Carter[3] and Philippos C. Patsalis[1,*]

[1]Department of Cytogenetics and Genomics, The Cyprus Institute of Neurology and Genetics, Nicosia 2370, Cyprus, [2]Laboratory of Cytogenetics, First Pediatric Department, Hippocration General Hospital, Thessaloniki 54642, Greece, [3]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK, [4]Department of Medical Genetics, University of Athens, St Sophia Children's Hospital, Athens 11527, Greece, [5]Centre for Human Genetics, University Hospital, Catholic University of Leuven, 3000 Leuven, Belgium, [6]Pediatric Endocrine Unit, Makarios III Hospital, Nicosia 1474, Cyprus, [7]Institute of Molecular and Cell Biology, University of Tartu, Tartu 51010, Estonia and [8]Medical School, University of Ioannina, Ioannina 45110, Greece

Received January 5, 2011; Revised and Accepted February 18, 2011

The recently described DNA replication-based mechanisms of fork stalling and template switching (FoSTeS) and microhomology-mediated break-induced replication (MMBIR) were previously shown to catalyze complex exonic, genic and genomic rearrangements. By analyzing a large number of isochromosomes of the long arm of chromosome X (i(Xq)), using whole-genome tiling path array comparative genomic hybridization (aCGH), ultra-high resolution targeted aCGH and sequencing, we provide evidence that the FoSTeS and MMBIR mechanisms can generate large-scale gross chromosomal rearrangements leading to the deletion and duplication of entire chromosome arms, thus suggesting an important role for DNA replication-based mechanisms in both the development of genomic disorders and cancer. Furthermore, we elucidate the mechanisms of dicentric i(Xq) (idic(Xq)) formation and show that most idic(Xq) chromosomes result from non-allelic homologous recombination between palindromic low copy repeats and highly homologous palindromic LINE elements. We also show that non-recurrent-breakpoint idic(Xq) chromosomes have microhomology-associated breakpoint junctions and are likely catalyzed by microhomology-mediated replication-dependent recombination mechanisms such as FoSTeS and MMBIR. Finally, we stress the role of the proximal Xp region as a chromosomal rearrangement hotspot.

INTRODUCTION

Mammalian autosomal chromosomes undergo homologous recombination during meiosis, ensuring proper chromosomal segregation and providing the means for increasing genetic diversity and for repairing chromosomal damage. While for autosomes, a homologous chromosome is always available for pairing during meiosis, human sex chromosomes face particular challenges. The lack of a homologous Y chromosome in male meiosis has led to the progressive decay of the Y chromosome and the evolution of massive palindromic low copy repeats (LCRs or segmental duplications) in the male-specific region of the Y chromosome (MSY), which harbor critical genes for spermatogenesis (1). Even though these palindromes predate the human-chimpanzee divergence, they exhibit sequence identity >99% suggesting that they are undergoing extensive sequence homogenization by gene conversion, thus providing a mechanism for maintaining the

*To whom correspondence should be addressed at: Department of Cytogenetics and Genomics, The Cyprus Institute of Neurology and Genetics, 6, International Airport Avenue, Nicosia 2370, Cyprus. Tel: +357 22392600; Fax: +357 22358237; Email: patsalis@cing.ac.cy

integrity of critical spermatogenesis genes in the absence of homologous recombination (1). The MSY palindromes also predispose to chromosomal rearrangements since they can act as substrates for non-allelic homologous recombination (NAHR), resulting in non-reciprocal exchanges and the formation of isodicentric Y chromosomes which are responsible for a significant proportion of cases with spermatogenic failure, Turner syndrome and sex reversal (2).

In the case of the X chromosome, while a homologue is available for pairing and recombination during female meiosis, the same challenges to those faced by the Y chromosome are present in male meiosis where only the pseudoautosomal regions of the mammalian sex chromosomes can pair and recombine. Also, while the pericentromeric region of the human X chromosome is not enriched for segmental duplications (3,4), in analogy to the MSY, the proximal Xp region is highly enriched for large and highly homologous palindromic segmental duplications which harbor genes that are expressed mainly or exclusively in testes (5). The proximal Xp palindromes, like the MSY palindromes, predate the human-chimpanzee divergence and yet exhibit very high arm-to-arm sequence identity (5). The structural similarities between the MSY and the proximal Xp palindromes and the occurrence of analogous non-pathogenic recombination processes in both the MSY and proximal Xp palindromes, as evidenced by the extensive arm-to-arm sequence homogenization, prompted us to hypothesize that analogous processes of pathogenic copy number change are likely to operate in proximal Xp. In order to test this hypothesis and investigate whether the proximal Xp palindromes catalyze the formation of isochromosomes of the long arm of chromosome X (i(Xq)), we analyzed a large number of Turner syndrome patients carrying i(Xq) chromosomes. Most i(Xq) chromosomes have breakpoints in proximal Xp, consist of mirror-image Xq arms and lack most or all of the Xp arm. The i(Xq) is the most common human isochromosome and the most frequent structural abnormality in Turner syndrome (6,7). Previous studies have established that i(Xq) chromosomes can be of either paternal or maternal origin (8,9) with approximately equal frequencies and that i(Xq) formation is not associated with increased parental age (8). It has also been shown that the majority of i(Xq) are dicentric isochromosomes and consist of identical long arms, suggesting that they originate from a single X chromosome both in the male and female germline, and that most of them do not arise by centromere misdivision (8–10). The mechanism of i(Xq) formation remains unclear.

The detailed analysis of 34 i(Xq) chromosomes allowed us to gain insights into the mechanisms of i(Xq) formation and identify specific genomic architectural elements which catalyze the formation of i(Xq) chromosomes via recombination and replication-based mechanisms.

RESULTS

Characterization of i(Xq) chromosomes by STR analysis

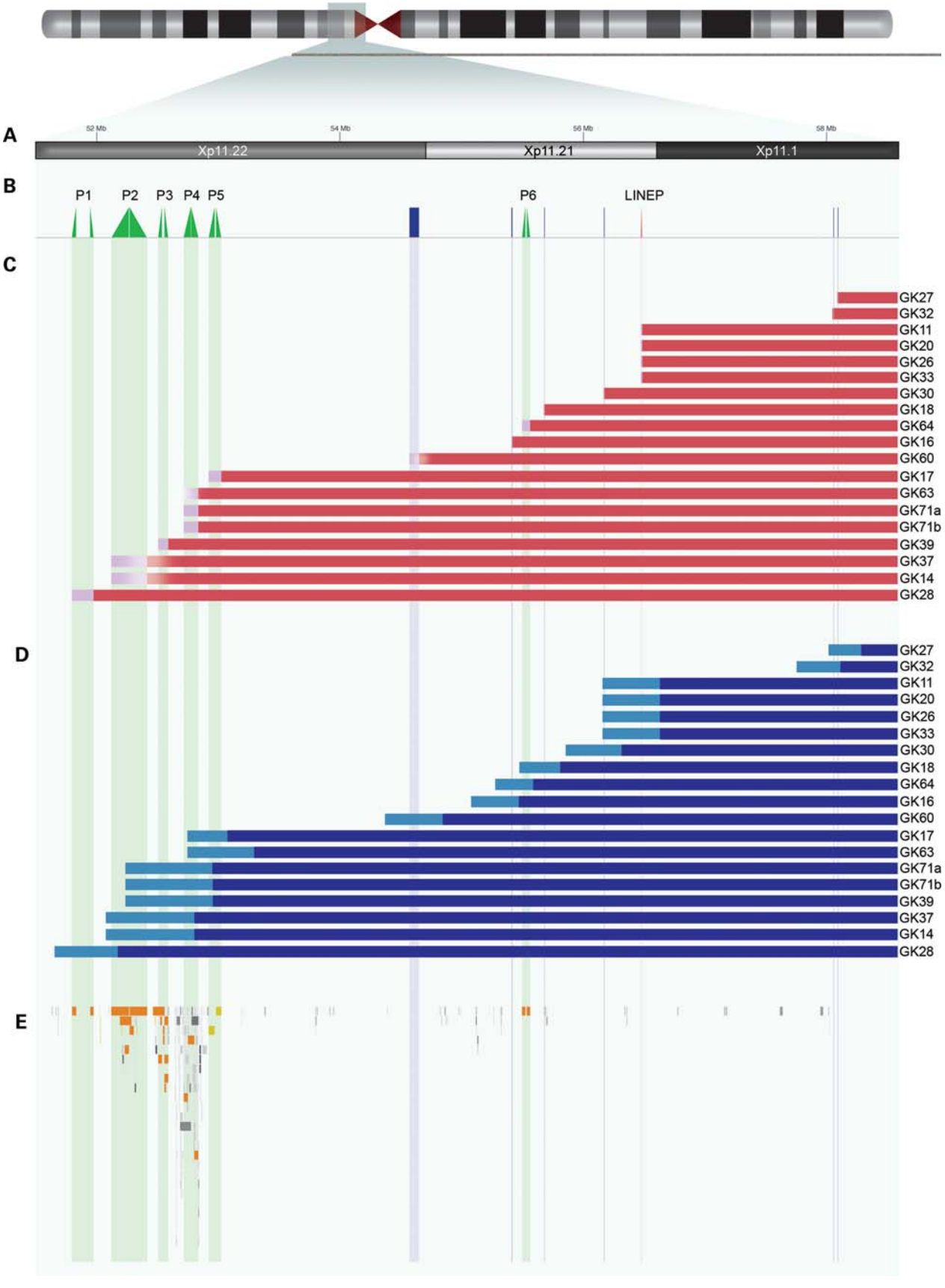
We first investigated whether the i(Xq) chromosomes are formed by an interchromosomal or intrachromosomal mechanism, by undertaking the amplification of a series of highly polymorphic short tandem repeat markers (STRs) spanning the X chromosome (Supplementary Material, Table S1). Homozygosity was observed at all tested Xp STRs. This is consistent with the deletion of almost the entire Xp arm on the isochromosome. In all but two cases, one or two different alleles were observed for the Xq markers. These results are consistent with an intrachromosomal mechanism involving sister chromatids. Interchromosomal recombination between homologous X chromosomes would lead to the presence of three alleles for most Xq markers, one derived from the long arm of the normal X chromosome and two different ones from the two different Xq arms of the i(Xq). In two cases, only one allele was detected at all tested loci due to very high mosaicism for a 45,X cell line. In another case, three alleles were observed at one Xq marker and in an additional case three alleles were detected at two markers. These findings suggest that the i(Xq) in this study consist of two identical arms, which are derived from sister chromatids. In the two cases where triallelic patterns were observed for one and two STRs, respectively, the possibility that the isochromosome originates from two homologous chromosomes cannot be excluded. However, since in these cases one or two alleles were observed at the majority of Xq STRs, the observed triallelic pattern of these markers is more likely caused by rare variants and/or allelic recombination at these loci.

High-resolution characterization of i(Xq) chromosomes by whole-genome tiling path array comparative genomic hybridization

Conventional cytogenetic methods such as karyotyping and fluorescence in situ hybridization (FISH) cannot efficiently distinguish monocentric i(Xq) from dicentric i(Xq) which harbor two centromeres in very close proximity. In order to unambiguously distinguish monocentric i(Xq) from dicentric i(Xq), we characterized the breakpoints of all 34 i(Xq) by using a whole-genome tiling path array (WGTPA) (11,12). We were able to identify the breakpoints of all 34 cases within two to four consecutive array clones (breakpoint resolution ~200–500 Kb) (Fig. 1D).

In 15 out of the 34 i(Xq) cases, the array clones corresponding to the short arm of chromosome X revealed a deletion of the entire euchromatic portion of the short arm, and all the long arm clones revealed a duplication of the entire long arm. Representative X-chromosome WGTPA data from five cases are shown in Supplementary Material, Figure S1A.

Figure 1. Microarray-determined breakpoints of idic(Xq) in relation to regional genomic architecture. (**A**) Expanded view of the 7 Mb proximal Xp region in which all the idic(Xq) breakpoints were localized. (**B**) Breakpoint annotation in relation to the underlying genomic architecture. The arms of the LCR palindromes i(Xq)-P1 to i(Xq)-P6 are denoted as paired green triangles (P1–P6). The LINE, L1 palindrome i(Xq)-LINEP is shown as a red triangle pair (LINEP). Breakpoint sequences in regions of no extended homology are denoted as blue rectangles. (**C**) High-resolution custom oligo aCGH breakpoints of idic(Xq). Red bars denote proximal Xp sequences that are duplicated on the idic(Xq). Lighter red bars denote breakpoint intervals. Color gradients in four cases indicate breakpoint uncertainty (see text). (**D**) WGTPA breakpoints of idic(Xq). Blue bars denote proximal Xp sequences that are duplicated on the idic(Xq). Lighter blue bars denote breakpoint intervals. (**E**) Segmental duplications in proximal Xp. The large, complex LCR cluster which includes i(Xq)-P1 to i(Xq)-P5 can be seen between ~52 and 53 Mb.



These 15 i(Xq) cases have breakpoints within centromeric heterochromatin and are thus defined as cytogenetically monocentric.

The remaining 19 of the 34 i(Xq) cases had WGTPA-determined breakpoints in euchromatic sequences within a 7 Mb region of proximal Xp and are thus defined as cytogenetically dicentric (idic(Xq)). The breakpoints of 8 of these 19 cases were localized within a complex cluster of large, highly homologous, direct and inverted LCRs (ChrX: ~52–53 Mb). The breakpoints of the remaining 11 cases were localized more proximally in an LCR-poor part of this 7 Mb region where only one large and highly homologous LCR palindrome is present (Fig. 1A, B, D, E).

Refinement of cytogenetically monocentric i(Xq) breakpoints by real-time polymerase chain reaction

In order to refine the WGTPA-determined breakpoints of the cytogenetically monocentric i(Xq) and determine whether they are structurally monocentric or structurally dicentric, primers were designed and quantitative real-time polymerase chain reaction (RT-PCR) was undertaken in the last unique sequence on proximal Xp, which lies deep into the centromeric heterochromatin, 606 Kb from *ZXDA*, the most proximal gene on Xp. This sequence lies within monomeric DXZ1 sequences close to the DXZ1 Array Junction which separates monomeric DXZ1 DNA from the higher order, DXZ1 array (13,14) (Supplementary Material, Fig. S1B). The functional centromere of chromosome X was previously delimited within the DXZ1 array (14). In most cases, we were able to determine whether the breakpoints of these i(Xq) chromosomes occur within monomeric alpha satellite DNA or more proximally, potentially within the DXZ1 array which is necessary for proper centromere function. In four cases, the breakpoints were localized within monomeric alpha satellite sequences, thus these i(Xq) are structurally dicentric. In three cases, we could not determine whether the breakpoints lie within monomeric alpha satellite sequences, or DXZ1 sequences due to very low mosaicism; however, these i(Xq) are apparently structurally dicentric, as evidenced by the high mosaicism for a 45,X cell line. Finally, in eight cases, the breakpoints were localized more proximally, and lie either within the most proximal part of the monomeric DXZ1 sequences, or within the DXZ1 array. Four of these eight cases are mosaic for a 45,X cell line, while the other four have a 46,X,i(Xq) karyotype. Considering that only structurally dicentric isochromosomes can behave as functionally dicentric (as opposed to structurally monocentric isochromosomes which can only be functionally monocentric) and give rise to a second 45,X cell line due to the loss of an unstable dicentric isochromosome, or due to isochromosome loss resulting from malsegregation at the time of i(Xq) formation (15), we hypothesize that only four of these eight i(Xq) can be structurally dicentric.

Ultra-high-resolution characterization of idic(Xq) breakpoints by custom oligonucleotide array comparative genomic hybridization

The WGTPA comparative genomic hybridization (CGH) identification of a common 7 Mb breakpoint interval and the considerable overlap of many idic(Xq) breakpoints with large and highly homologous LCR palindromes prompted us to investigate the potential involvement of specific palindromes in idic(Xq) formation and determine the fine-scale structure of the idic(Xq) breakpoint junctions. Two custom high-resolution (mean resolution: ~180 bp) and ultra-high-resolution (mean resolution: ~20 bp) microarrays targeting the 7 Mb breakpoint region and individual WGTPA-determined breakpoints were designed. While typical microarray designs do not include probes in repetitive genomic regions, in this study we followed an alternative approach carefully designing array probes which provide comprehensive coverage of both unique sequences in proximal Xp and also of palindrome arms and spacers.

We were able to map the breakpoints of nine idic(Xq) cases to the junctions of highly homologous large LCR palindromes defined as i(Xq)-P1 to i(Xq)-P6 (Fig. 1B and C, Supplementary Material, Figs S2 and S3). Almost every large palindrome in the region was associated with at least one idic(Xq) breakpoint and the two largest palindromes with arm homologies >99.5% were targeted more than once. The breakpoints of two idic(Xq) were localized to i(Xq)-P2. This is the biggest palindrome in the region and consists of 142 Kb arms that have >99.9% sequence identity and are separated by an 8 Kb spacer. The breakpoints of the idic(Xq) from the twin patients along with the breakpoint of another unrelated case were localized to i(Xq)-P4. This is the second biggest palindrome in the region. It consists of 60 Kb arms that share >99.5% sequence identity. The two arms are separated by a very small 0.2 Kb spacer. Besides their very large size, the i(Xq)-P2 and i(Xq)-P4 palindromes are characterized by additional complexity. The proximal junction of the massive i(Xq)-P2 palindrome is flanked by one of the few remaining gaps in the human reference genome, and the arms and spacers of both palindromes are partially duplicated and organized in complex duplicon clusters in this region and in other parts of the X chromosome. The breakpoint of another case was localized in i(Xq)-P3. This palindrome consists of 29 Kb arms that have >99.8% sequence identity and are separated by a 22 Kb spacer. Additional complexity is also present at this palindrome as both arms and the spacer are duplicated in multiple positions in this region, and the palindrome in its entirety comprises the central section of i(Xq)-P2. The breakpoints of cases GK28 and GK17 were localized to the junctions of i(Xq)-P1 and i(Xq)-P5, respectively. These two palindromes flank the complex LCR cluster. The i(Xq)-P1 flanks the LCR cluster distally and consists of 36 Kb arms that have sequence identity >99.8% and are separated by a 100 Kb spacer. This palindrome does not exhibit additional complexity and has an unusually large spacer compared with the other breakpoint-associated palindromes. The i(Xq)-P5 flanks the LCR cluster proximally and consists of 39 Kb arms that have 97.9% sequence identity and are separated by a 15 Kb spacer. Finally, the breakpoint of the last palindrome-associated idic(Xq) was localized more proximally at i(Xq)-P6. This is the most proximal large and highly homologous palindrome and consists of 27 Kb arms that have sequence identity >99.8% and are separated by a 13 Kb spacer. Overall, a preference for high sequence identity, large arm size and small spacer length was observed, as the palindromes that exhibited these characteristics were targeted more than once. The presence of unique sequences at the spacers of i(Xq)-P1, i(Xq)-P5 and

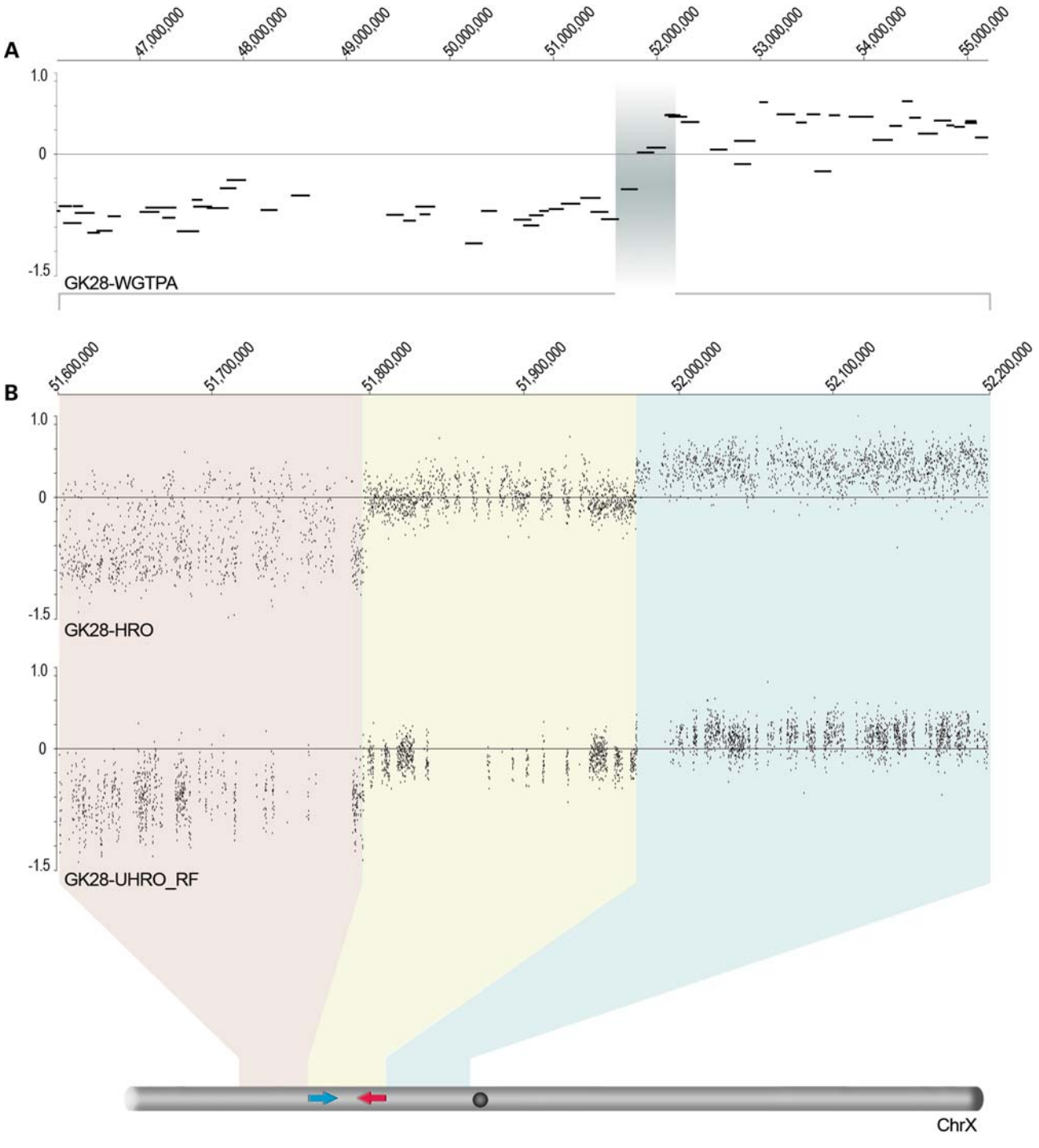


Figure 2. The fine-scale structure of palindrome-associated idic(Xq) breakpoints. (**A**) idic(Xq) breakpoint characterized by WGTPA CGH. Highlighted breakpoint clones exhibit intermediate ratios in relation to duplicated and deleted clones. (**B**) Expanded view of highlighted breakpoint region showing non-repeat-filtered aCGH data from the high-resolution oligo array (GK28-HRO) and repeat-filtered aCGH data from the ultra-high-resolution oligo array (GK28-UHRO_RF) in relation to i(Xq)-P1. Light-red highlighted probes represent sequences immediately distal to the left palindrome arm and are deleted. Light-blue highlighted probes represent sequences immediately proximal to the right palindrome arm and are duplicated. Light-yellow highlighted probes represent sequences corresponding to the palindrome arms and unique-sequence spacer and are present in one copy on the isochromosome. For clarity, chromosome X is not shown to scale.

i(Xq)-P6, and the absence of additional complexity at the arm sequences of these palindromes allowed the characterization of the palindrome-specific breakpoint sequences (Fig. 2, Supplementary Material, Figs S2 and S3). Sequences immediately distal to the breakpoint palindromes are deleted and sequences immediately proximal to the palindromes are duplicated on the isochromosomes. More importantly, as can be seen in Figure 2, Supplementary Material, Figures S2 and S3, the

palindrome arms and spacer are intact on the isochromosome (present in one copy). These results suggest that crossovers (COs) between palindrome arms are likely to be the initiating event giving rise to the idic(Xq) that have palindrome-associated breakpoints. The extensive complexity of i(Xq)-P2, i(Xq)-P3 and i(Xq)-P4, which are characterized by complex duplication patterns and a reference sequence gap, and also the high mosaicism for a 45,X cell line made such determinations not possible in cases GK14, GK37 and GK63.

In 4 of the 19 idic(Xq) cases, the breakpoints were localized within a LINE, L1 rich region. The high-resolution aCGH revealed the same breakpoint in all four cases. The breakpoint was localized to the junctions of a LINE, L1 palindrome defined as i(Xq)-LINEP (Fig. 1, Supplementary Material, Fig. S4). This palindrome consists of two almost identical ~5.5 Kb LINE, L1PA3 repeats arranged in an inverted orientation. The two repeats have >99% sequence identity and are separated by a very short spacer which consists of 30 bp of overlap between the two inverted L1s. This is the most recombinogenic structure we identified in this study and catalyzes the formation of four idic(Xq).

The breakpoints of the remaining 6 of the 19 idic(Xq) cases were localized in regions of no extended homology (Fig. 1, Supplementary Material, Fig. S5). The breakpoint of one of these cases could not be refined extensively due to very high mosaicism for a 45,X cell line. Another case had a breakpoint in a direct intrachromosomal segmental duplication. The breakpoints of the remaining four cases were localized in interspersed repeat-rich sequences.

Characterization of non-LCR associated idic(Xq) breakpoints by sequencing

Unique primers were designed and RT-PCR was undertaken in order to refine and further characterize the idic(Xq) chromosomes whose breakpoints were localized in regions of no extended homology. Based on the custom oligo aCGH and RT-PCR results, primers spanning the breakpoints were designed and breakpoint junctions were obtained by PCR in three cases and were cycle sequenced.

Microhomologies of 2–5 bp were identified at the breakpoints of all three cases (Fig. 3). The microhomologies are present in inverted orientation on opposite strands at the proximal and distal breakpoint junctions. A microhomology of 2 bp (GC) is present at the breakpoint of case GK16. The proximal junction is located in an SVA repeat and the distal junction is located in unique sequences (Supplementary Material, Fig. S5). The distal and proximal breakpoint junctions of this idic(Xq) are 3.69 Kb apart. Three base pairs of microhomology (GAG) are present at the breakpoint of case GK27. The proximal breakpoint junction of this case is located in a SINE, Alu repeat and the distal breakpoint junction is located in an LTR, ERVL repeat. The breakpoint junctions are 1.97 Kb apart. Finally, five bases of microhomology (TTTAT) are present at the breakpoint of case GK32. The two junctions are 6.82 Kb apart and both reside in LINE, L1 repeats. In all three cases, the chromosomal region between the proximal and distal junctions (defined as template switch region) is neither duplicated nor deleted, but present in one copy on the isochromosome. No additional complexity was detected at the breakpoints. Even though the breakpoints of the remaining two cases were thoroughly refined, we were unable to isolate them by PCR, which suggests that further breakpoint complexity might be present, thus rendering them refractory to PCR analysis.

DISCUSSION

In order to characterize the molecular basis of i(Xq) formation and investigate the role of the proximal Xp LCR palindromes in the formation of X-chromosome rearrangements, we analyzed i(Xq) chromosomes from a large number of Turner syndrome cases using high-resolution molecular methodologies including high- and ultra-high-resolution aCGH, and sequencing. We were able to map the breakpoints of 34 i(Xq) chromosomes, identify the molecular mechanisms which may be responsible for the formation of the idic(Xq) and show that most if not all dicentric i(Xq) are catalyzed by the human genome's underlying genomic architecture.

Intrachromosomal recombination between inverted alpha satellite repeats may be the prominent mechanism for the formation of cytogenetically monocentric i(Xq) chromosomes

The breakpoints of 15 out of 34 cases were localized in centromeric heterochromatin. Even though crossing over is suppressed near and within centromeres (16) in order to prevent non-disjunction and chromosome breakage (17,18), unequal sister-chromatin exchange and/or gene conversion are thought to underlie the concerted evolution of centromeric sequences (13,19,20).Topoisomerase IIa has been shown to cleave hairpins formed in human centromeric DNA (21). Such hairpin structures can theoretically form by inverted sequences, and the recent discovery of a small polymorphic inversion in chromosome X centromeric sequences (22) suggests that inversions within centromeric DNA may not be a rare phenomenon. Also, substantial double-strand break (DSB) formation was documented near centromeres in budding yeast (23), and a recent study has found evidence of widespread gene conversion in maize centromeres where crossing over is also suppressed (24). It appears that DSB formation is frequent at centromeres and that recombination in centromeric sequences is regulated at the level of DNA repair.

The i(Xq) chromosomes which have breakpoints in centromeric sequences can theoretically result from CO resolution of NAHR events between inverted centromeric sequences on sister chromatids. When recombination occurs more proximally within the DXZ1 array, the resulting isochromosome would carry a truncated DXZ1 array and would be functionally monocentric. A more distal NAHR event within monomeric alpha satellite DNA would give rise to a structurally dicentric i(Xq), carrying two DXZ1 arrays. Depending on the degree of coordination between the two DXZ1 arrays and the effect/extent of the poorly understood phenomenon of centromere inactivation, such a structurally dicentric isochromosome could be either functionally dicentric, or functionally monocentric. Our data are in agreement with previous reports (8,10) and suggest that the vast majority of

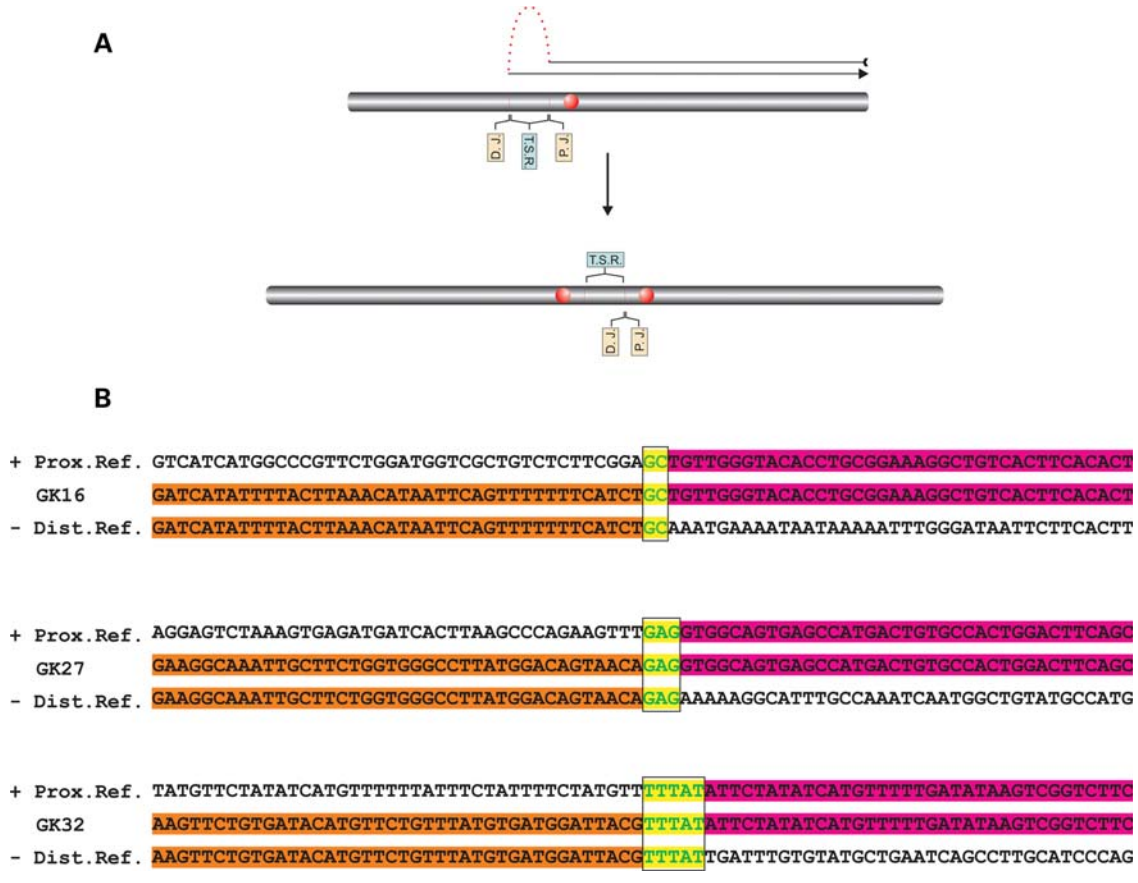


Figure 3. Microhomology-associated breakpoints. (**A**) Replication of the entire Xq arm and a small part of proximal Xp, represented as a solid horizontal line above the chromosome X illustration, followed by a template switch (dotted line) more distally and replication in the reverse direction to the end of the chromosome can catalyze the formation of idic(Xq) that have non-recurrent breakpoints. Proximal and distal junctions (P.J. and D.J.) represent the sites of replication disruption and resumption, respectively. The template switch region (T.S.R.) is intact on the idic(Xq) (present in one copy). Microhomologies are illustrated as vertical red lines. Alternatively, the proximal-distal template switch could occur in the reverse order. (**B**) Three sequenced breakpoint junctions. Plus strand proximal reference sequence is shown at the top. Minus strand distal reference sequence is shown at the bottom. Breakpoint sequence is shown in the middle. Microhomology bases (green letters) are boxed. All sequences are shown 5′–3′.

i(Xq) chromosomes are structurally dicentric. Alternatively, isochromosome formation by the classical model of centromere misdivision (25) or via a replication-based mechanism cannot be excluded.

NAHR between palindromic sequences catalyzes the formation of recurrent-breakpoint idic(Xq) chromosomes

The breakpoints of nine idic(Xq) chromosomes were localized to the junctions of large and highly homologous palindromes which in analogy to the MSY palindromes on the Y chromosome are enriched for genes that are predominantly or exclusively expressed in testes. Also, the breakpoints of four idic(Xq) were localized to a LINE, L1 palindrome. Based on these findings, and the fact that long palindromic sequences are recombinogenic structures which can extrude into hairpins or cruciforms and induce the formation of DSBs via the action of structure-specific nucleases, we hypothesize that recombination between the highly homologous arms of proximal Xp palindromes is responsible for the formation of the palindrome-associated idic(Xq) and propose a model which offers a mechanistic interpretation of how these rearranged chromosomes arise (Supplementary Material, Fig. S6). We propose that a DSB within one arm of a palindrome can initiate a NAHR event in the same manner that DSBs initiate homologous recombination during meiosis. DSB formation followed by 5′ to 3′ resection and strand invasion of homologous sequences on the opposing palindrome arm, on the sister chromatid and subsequent DNA repair synthesis, ligation and CO resolution of the resulting double Holliday junctions can lead to the formation of an isodicentric chromosome and an acentric fragment. The acentric fragment is eventually lost due to the lack of a centromere, and the isodicentric chromosome is stabilized through inactivation of one of the two centromeres. Incomplete centromere inactivation is a frequent phenomenon, and results in mosaic 45,X/46,X,i(Xq) karyotypes.

A number of testable predictions stem from the proposed model. First, the resulting isochromosome should carry a duplication of the entire Xq arm. Second, if the NAHR event involves two sister chromatids, the two Xq arms should be identical. Third, the Xp sequences proximal to the palindrome should be duplicated, and the Xp sequences distal to the palindrome should be deleted. Finally, the arms

and the spacer of the breakpoint-associated palindrome should be intact on the isochromosome (no deletions or duplications of palindrome sequences). As shown above, all four predictions were experimentally tested and the results support the proposed model.

Previous studies have localized the breakpoints of isochromosome 17q within large palindromic LCRs on 17p11 (26) and the breakpoints of the isodicentric Y chromosome within MSY palindromes (2), suggesting that pericentromeric LCR palindromes play an important role in isochromosome formation. A resent study (27) identified breakpoints of eight idic(Xq) cases within proximal Xp and the breakpoints of seven of these cases were localized within regions encompassing four of the six i(Xq) palindromes identified here. Even though palindrome probe coverage was not provided in this study and thus, the copy number state of sequences within the palindrome arms and spacers could not be determined, the emerging breakpoint landscape for the NAHR-catalyzed i(Xq) chromosomes appears to be characterized by highly recurrent breakpoints involving most, if not all, of the highly homologous palindromic LCRs in proximal Xp and a single highly homologous LINE, L1 palindrome.

Microhomology-mediated replication-dependent recombination, or non-homologous end joining catalyzes the formation of non-recurrent-breakpoint idic(Xq) chromosomes

The lack of extended homology at the breakpoints of six idic(Xq) precludes the possibility of being generated by a homologous recombination repair mechanism, since extensive homology of up to 300 bp is needed for homologous recombination in humans (28). Even though non-homologous end joining (NHEJ) cannot be excluded as a potential mechanism of formation, the microhomologies at the breakpoints of three cases, along with perfect sequence preservation of both the proximal and distal breakpoint-junction sequences, suggest that the involvement of a replication-based mechanism is more likely. The absence of the common 'information scar' (29) left at the breakpoints of many NHEJ-repaired sequences implies that no end processing took place. For these breakpoints to have been generated by NHEJ, DSB formation at different positions in close proximity, at sites of perfect microhomology, on the two sister chromatids, would be needed in all three cases. Such an event would then need to be repaired distally and not locally by joining together opposite strand sequences from the two DSB sites. These findings can be more readily and economically explained by a microhomology-mediated replication-dependent recombination mechanism (30) such as serial replication slippage (SRS) (31–33), break-induced SRS (34), fork stalling and template switching (FoSTeS) or microhomology-mediated break-induced replication (MMBIR) (35,36). The presence of breakpoint microhomology, the relatively close proximity of the breakpoint junctions and the nearby presence of complex LCRs are the hallmarks of the recently described FoSTeS (35), and MMBIR mechanisms (36). FoSTeS provides the conceptual and mechanistic framework to explain the formation of non-recurrent and complex rearrangements which cannot be explained by NAHR or NHEJ. FoSTeS events tend to occur in regions where complex genomic architecture is thought to impede replication fork progression. Disengagement of the lagging strand and annealing at a site of inverted microhomology at a nearby replication fork moving in the opposite direction (Supplementary Material, Fig. S7A), or at any nearby single-stranded sequence having inverted microhomology with the free 3′-end, followed by resumption of replication to the end of the chromosome (FoSTeS X 1) is consistent with our findings in all three cases whose breakpoints were sequenced.

MMBIR repair is also consistent with our findings. MMBIR incorporates further molecular mechanistic details based on the break-induced replication (BIR) repair mechanism and mediates the repair of single-ended DSBs arising from replication fork collapse in cells under stress (36,37). In such an event, if a replication fork encounters a DNA nick, a single-ended DSB could be generated (Supplementary Material, Fig. S7Bb, Bc, Bd). Subsequent 5′ to 3′ resection could allow the 3′-end to anneal to a nearby single-stranded sequence which shares inverted microhomology with the 3′-end, and prime DNA synthesis in the opposite direction (Supplementary Material, Fig. S7Be and Bf). Establishment of a replication fork with both leading and lagging strand synthesis (Supplementary Material, Fig. S7Bg) would result in the formation of an idic(Xq) (Supplementary Material, Fig. S7Bh).

Five of the six breakpoint junctions in the three sequenced cases reside within interspersed repeats. Repetitive sequences have a higher propensity to form non-B DNA structures and can promote the formation of DNA nicks, DSBs or stall replication forks inducing chromosomal rearrangements (38,39). Also, at both the proximal and distal breakpoint junctions of the three sequenced idic(Xq) breakpoints, sequences which can form secondary structures when in single stranded form were identified (Supplementary Material, Figs S8 and S9). Such sequences can potentially form hairpins or fold into more complex secondary structures on the lagging strand template during replication, where ~200 bp of single-stranded DNA is needed for Okazaki fragment priming (40). Whether these structures form *in vivo* and whether they can cause replication forks to stall or collapse remains unknown. However, it is interesting to note that in all six breakpoint junctions where alternative DNA structures can potentially form, the microhomology junctions are located within terminal or internal hairpin loops/bulges near single-/double-strand transition regions (Supplementary Material, Figs S8 and S9). Such structures can be cleaved by structure-specific nucleases such as the Artemis:DNA-PKcs complex (38,41). Also, since the microhomology junctions reside within single-stranded regions, they can potentially be accessed by collapsed replication forks for annealing via MMBIR. Nevertheless, since predicted secondary structures are prevalent throughout the genome, the actual significance of these observations cannot be currently determined.

Further support for the involvement of a replication-based mechanism in the formation of the non-recurrent breakpoint idic(Xq) comes from two recent studies showing that dicentric and acentric chromosomes in yeast are readily formed by a replication-based mechanism that involves template switching and may not require DSBs (42,43). Also, recently, post-zygotically

formed human isochromosomes were detected in cleavage-stage embryos (44).

The identification of these non-homologous recombination-catalyzed isochromosomes suggests that this region is susceptible to both recombination-based and replication-based rearrangements. While replication-based mechanisms such as FoSTeS and MMBIR have been shown to catalyze the formation of other chromosomal rearrangements (35,45–47), this is, to our knowledge, the first report of the involvement of replication-based mechanisms in the formation of human isochromosomes. Such replication-based mechanisms can also provide a potential explanation for the genesis of the complex segmental duplications found in proximal Xp. Theoretically, initial direct or inverted duplications could readily form via replication-based mechanisms which do not require extended homology. These initial 'seeder' LCRs could then serially propagate and disperse through repeated NAHR-catalyzed duplication events giving rise to complex segmental duplication clusters such as the ones seen in proximal Xp.

Human proximal Xp: a chromosomal rearrangement hotspot

A large number of diverse chromosomal rearrangements have been previously localized in proximal Xp. Approximately 30% of the non-syndromic X-linked mental retardation genes are located in Xp11 (48,49) and the recently described recurrent dup(X)(p11.22-p11.23) which is associated with speech delay and electroencephalographic anomalies (50) exhibits breakpoints that overlap with i(Xq)-P4. This palindrome catalyzes the formation of idic(Xq) chromosomes in three cases in the present study. Also, breakpoints of X-chromosome translocations and more complex rearrangements were previously mapped within Xp11.2 (51,52). This region is also associated with cancer. A variant associated with prostate cancer risk was recently identified in proximal Xp (53), and the reciprocal translocations t(X;18) have X-chromosome breakpoints in i(Xq)-P4 and other large LCRs in proximal Xp which harbor 'cancer antigen' genes (54). Polymorphic inversions involving four of the six i(Xq) palindromes (Supplementary Material, Fig. S10) and overall a significant enrichment for polymorphic inversions in proximal Xp were previously identified (55). These data, along with the data presented in this study, suggest that both interchromatid and intrachromatid recombination involving the proximal Xp LCR palindromes is a frequent phenomenon.

Classical meiotic homologous recombination can be alternatively resolved via non-crossover (NCO) or CO pathways. NCO resolution of Holliday junctions results in gene conversion, while CO resolution results in crossing over (56). Considering the mechanistic analogies between meiotic homologous recombination and NAHR, we postulate that NCO resolution of recombination events between opposing palindrome arms on the same chromatid (Supplementary Material, Fig. S11Aa) or on sister chromatids (Supplementary Material, Fig. S11Ba) would lead to gene conversion. Alternatively, CO resolution of intrachromatid recombination between opposing palindrome arms (Supplementary Material, Fig. S11Ab) would lead to inversions, while CO resolution of interchromatid recombination between opposing palindrome arms would lead to the formation of dicentric isochromosomes (Supplementary Material, Fig. S11Bb). We propose that the extensive arm-to-arm sequence homogenization, the identified polymorphic inversions and the isochromosome formation involving the palindromic sequences in proximal Xp constitute alternative outcomes of recombination events between palindromic sequences.

The unique biology of the X chromosome and the diverse genomic architecture of human proximal Xp, which is characterized by a high concentration of recombinogenic structures, render this region a hotspot for both recombination-based and evidently replication-based rearrangements which provide significant insights into the genesis of not only LCRs, but also of complex constitutional and somatic rearrangements, such as the ones seen in genomic disorders and many cancers. The human proximal Xp region offers an insightful look into the evolutionary and mutational processes which shaped the human genome.

MATERIALS AND METHODS

Study cases

The i(Xq) cases in this study consisted of 32 unrelated Turner syndrome patients that had either a 46,X,i(Xq) karyotype or a 45,X/46,X,i(Xq) karyotype exhibiting various degrees of mosaicism. The study also included two identical twins with a 45,X/46,X,i(Xq) karyotype.

Informed consent was obtained from all subjects. The screening protocols and study design were approved by the review boards of all institutes involved in the study and by the Cyprus National Bioethics Committee. DNA was extracted from peripheral blood using the Gentra Puregene genomic DNA purification kit (Qiagen, Valencia, CA, USA) and, where material was available, cultures were setup and karyotyping was repeated using standard techniques in order to confirm the presence of the abnormal chromosome.

STR analysis

Quantitative fluorescent polymerase chain reaction amplification was performed as previously described (57). Three markers were derived from the short arm and five markers from the long arm of chromosome X. The identity and cytogenetic position of these markers are shown in Supplementary Material, Table S1.

Whole-genome aCGH

The initial screening of all cases was carried out using a WGTPA consisting of 26 574 BAC clones covering 93.7% of the euchromatic sequence of the human genome (11,12). In summary, 150 ng of test and reference DNA (male or female) was differentially labeled using the Bioprime Labeling Kit (Invitrogen, Carlsbad, CA, USA) and hybridized on the arrays using an automated slide processor (HS.4800, Tecan Inc., Mannedorf, Switzerland). Two hybridizations were run for each sample in a dye-swap mode. The arrays were scanned using the Agilent DNA microarray scanner (Agilent Technologies Inc., Santa Clara, CA, USA). Fluorescent

intensities were extracted, dye-swap experiments were fused and ratios were calculated using the BlueFuse software (BlueGnome Ltd, Cambridge, UK). Aberration calling was carried out using the BlueFuse software and custom Perl scripts (11).

FISH analysis

FISH validation of the WGTPA results was carried out in 11 i(Xq) cases in order to empirically assess the accuracy of the array results before proceeding with the design of targeted high-resolution oligonucleotide arrays. In summary, the WGTPA-determined breakpoint BAC clones were isolated from the X-chromosome tiling path array library. Probes were labeled using the Abbott Nick Translation Kit and Abbott Orange dUTP (Abbott Laboratories, Abbott Park, IL, USA). Hybridizations were carried out according to the manufacturer's recommended protocols.

High-resolution and ultra-high-resolution targeted aCGH

Once the breakpoints of the isochromosomes were determined using the BAC WGTPA, custom high-resolution and ultra-high-resolution oligonucleotide arrays (Agilent Technologies Inc.; NimbleGen Systems Inc., Madison, WI, USA) were designed in order to refine the breakpoint-junction determination. The NimbleGen array (custom 385K) was a non-segmental duplication-masked, non-repeat-masked tiling design covering the proximal region of Xp in which all the isochromosome breakpoints were localized by the WGTPA. The array consisted of 345 782 probes tiling the region ChrX: 51,000,131–58,597,616 in an unbiased manner (mean resolution: ~20 bp). The Agilent array (custom 44 K) consisted of 3723 autosomal probes, 75 chromosome Y probes, 39 089 probes covering the region ChrX: 51,373,761–58,598,658 (mean resolution: ~180 bp) and 215 backbone probes covering the rest of the X chromosome in low resolution. The autosomal probes were used for normalization. The chromosome X backbone probes and chromosome Y probes were used as controls to assess the hybridization efficiency. The custom 44 K array was also a non-segmental duplication-masked, non-repeat-masked design and was carefully enriched for probes in subregions of proximal Xp. These subregions were identified as isochromosome breakpoint junctions by the WGTPA. The two microarray designs were not subjected to repeat masking or segmental-duplication masking in order to investigate the potential involvement of both segmental duplications and interspersed repeats in the formation of the isochromosomes. LCR and repeat masking of the data was applied during data analysis where necessary. The probes on the custom 44 K array were assigned confidence scores based on GC content, Tm, sequence complexity, hairpin ΔG and homology with the rest of the genome (Agilent Technologies Inc.). Probe scores were used during data analysis to filter out poor-performing probes. Hybridizations and analysis using the custom 385 K array were performed by NimbleGen. The cases were also screened in-house using the custom 44 K array following the Agilent CGH protocol with the following modifications. Test and reference DNA (500 ng each) were differentially labeled overnight using the Bioprime Labeling Kit (Invitrogen). Hybridization, scanning and feature extraction were carried out according to Agilent's recommended protocols. Self–self hybridizations were carried out to empirically assess probe performance. Normalization and data analysis were carried out using custom scripts. Saturated probes and low-quality probes were removed from the data based on the results of the self–self control hybridizations. Repeat masking was carried out using the RepMask 3.2.7 track (58) from the UCSC Genome Browser March 2006 assembly and custom perl scripts. Probe filtering was performed using probe confidence scores.

RT-PCR analysis

In those cases where the array-determined breakpoints were mapped in alpha satellite repeats, or in clusters of interspersed repeats, array-based confirmation was not possible. In these cases, unique primers were designed and RT-PCR was performed to confirm and refine the array-determined breakpoint junctions. RT-PCR primers were designed using Primer3 (59) and were BLASTed (60) against the reference genome. RT-PCR was carried out using the SYBR Green PCR Master mix (Applied Biosystems, Foster City, CA, USA) and reactions were run on the ABI 7900 Real-Time PCR System (Applied Biosystems).

Sequencing and sequence analysis

In three cases, we were able to PCR amplify the breakpoint junctions by using outward-facing primers. PCR reactions were carried out using the illustra r*Taq* DNA Polymerase (GE Healthcare UK Limited, Little Chalfont, UK). Sequencing was carried out using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). Sequencing reactions were run on a 3130*xl* Genetic Analyzer (Applied Biosystems). DNA sequences were analyzed by comparing them to the reference genome using BLAT (61). Secondary structure predictions were carried out using mfold (62). Two hundred base pairs flanking the breakpoint microhomology junctions were analyzed for cases GK16 and GK27 and 1Kb flanking the breakpoint junctions of the idic(Xq) from case GK32. Sequences were analyzed using default mfold parameters. Untangle modes were used in order to avoid overlaps.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank the patients and their families for their participation and Dr Heike Fiegler for assisting with experiments.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the Cyprus Research Promotion Foundation (EPYEΞ 0406). A.K. received support from the Estonian Government (SF0180027s10), (ETF 7617).

REFERENCES

1. Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T. *et al.* (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, **423**, 825–837.
2. Lange, J., Skaletsky, H., van Daalen, S.K., Embry, S.L., Korver, C.M., Brown, L.G., Oates, R.D., Silber, S., Repping, S. and Page, D.C. (2009) Isodicentric Y chromosomes and sex disorders as byproducts of homologous recombination that maintains palindromes. *Cell*, **138**, 855–869.
3. She, X., Horvath, J.E., Jiang, Z., Liu, G., Furey, T.S., Christ, L., Clark, R., Graves, T., Gulden, C.L., Alkan, C. *et al.* (2004) The structure and evolution of centromeric transition regions within the human genome. *Nature*, **430**, 857–864.
4. Ross, M.T., Grafham, D.V., Coffey, A.J., Scherer, S., McLay, K., Muzny, D., Platzer, M., Howell, G.R., Burrows, C., Bird, C.P. *et al.* (2005) The DNA sequence of the human X chromosome. *Nature*, **434**, 325–337.
5. Warburton, P.E., Giordano, J., Cheung, F., Gelfand, Y. and Benson, G. (2004) Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. *Genome Res.*, **14**, 1861–1869.
6. Hook, E.B. and Warburton, D. (1983) The distribution of chromosomal genotypes associated with Turner's syndrome: livebirth prevalence rates and evidence for diminished fetal mortality and severity in genotypes associated with structural X abnormalities or mosaicism. *Hum. Genet.*, **64**, 24–27.
7. Palmer, C.G. and Reichmann, A. (1976) Chromosomal and clinical findings in 110 females with Turner syndrome. *Hum. Genet.*, **35**, 35–49.
8. James, R.S., Dalton, P., Gustashaw, K., Wolff, D.J., Willard, H.F., Mitchell, C. and Jacobs, P.A. (1997) Molecular characterization of isochromosomes of Xq. *Ann. Hum. Genet.*, **61**, 485–490.
9. Lorda-Sanchez, I., Binkert, F., Maechler, M. and Schinzel, A. (1991) A molecular study of X isochromosomes: parental origin, centromeric structure, and mechanisms of formation. *Am. J. Hum. Genet.*, **49**, 1034–1040.
10. Wolff, D.J., Miller, A.P., Van Dyke, D.L., Schwartz, S. and Willard, H.F. (1996) Molecular definition of breakpoints associated with human Xq isochromosomes: implications for mechanisms of formation. *Am. J. Hum. Genet.*, **58**, 154–160.
11. Fiegler, H., Redon, R., Andrews, D., Scott, C., Andrews, R., Carder, C., Clark, R., Dovey, O., Ellis, P., Feuk, L. *et al.* (2006) Accurate and reliable high-throughput detection of copy number variation in the human genome. *Genome Res.*, **16**, 1566–1574.
12. Fiegler, H., Redon, R. and Carter, N.P. (2007) Construction and use of spotted large-insert clone DNA microarrays for the detection of genomic copy number changes. *Nat. Protoc.*, **2**, 577–587.
13. Schueler, M.G., Dunn, J.M., Bird, C.P., Ross, M.T., Viggiano, L., Rocchi, M., Willard, H.F. and Green, E.D. (2005) Progressive proximal expansion of the primate X chromosome centromere. *Proc. Natl Acad. Sci. USA*, **102**, 10563–10568.
14. Schueler, M.G., Higgins, A.W., Rudd, M.K., Gustashaw, K. and Willard, H.F. (2001) Genomic and genetic definition of a functional human centromere. *Science*, **294**, 109–115.
15. Sullivan, B.A. and Willard, H.F. (1998) Stable dicentric X chromosomes with two functional centromeres. *Nat. Genet.*, **20**, 227–228.
16. Mahtani, M.M. and Willard, H.F. (1998) Physical and genetic mapping of the human X chromosome centromere: repression of recombination. *Genome Res.*, **8**, 100–110.
17. Lamb, N.E., Sherman, S.L. and Hassold, T.J. (2005) Effect of meiotic recombination on the production of aneuploid gametes in humans. *Cytogenet. Genome Res.*, **111**, 250–255.
18. Talbert, P.B. and Henikoff, S. (2010) Centromeres convert but don't cross. *PLoS Biol.*, **8**, e1000326.
19. Pironon, N., Puechberty, J. and Roizes, G. (2010) Molecular and evolutionary characteristics of the fraction of human alpha satellite DNA associated with CENP-A at the centromeres of chromosomes 1, 5, 19, and 21. *BMC Genomics*, **11**, 195.
20. Schindelhauer, D. and Schwarz, T. (2002) Evidence for a fast, intrachromosomal conversion mechanism from mapping of nucleotide variants within a homogeneous alpha-satellite DNA array. *Genome Res.*, **12**, 1815–1826.
21. Jonstrup, A.T., Thomsen, T., Wang, Y., Knudsen, B.R., Koch, J. and Andersen, A.H. (2008) Hairpin structures formed by alpha satellite DNA of human centromeres are cleaved by human topoisomerase IIalpha. *Nucleic Acids Res.*, **36**, 6165–6174.
22. McKernan, K.J., Peckham, H.E., Costa, G.L., McLaughlin, S.F., Fu, Y., Tsung, E.F., Clouser, C.R., Duncan, C., Ichikawa, J.K., Lee, C.C. *et al.* (2009) Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res.*, **19**, 1527–1541.
23. Blitzblau, H.G., Bell, G.W., Rodriguez, J., Bell, S.P. and Hochwagen, A. (2007) Mapping of meiotic single-stranded DNA reveals double-stranded-break hotspots near centromeres and telomeres. *Curr. Biol.*, **17**, 2003–2012.
24. Shi, J., Wolf, S.E., Burke, J.M., Presting, G.G., Ross-Ibarra, J. and Dawe, R.K. (2010) Widespread gene conversion in centromere cores. *PLoS Biol.*, **8**, e1000327.
25. de la Chapelle, A. (1982) How do human isochromosomes arise? *Cancer Genet. Cytogenet.*, **5**, 173–179.
26. Barbouti, A., Stankiewicz, P., Nusbaum, C., Cuomo, C., Cook, A., Hoglund, M., Johansson, B., Hagemeijer, A., Park, S.S., Mitelman, F. *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.
27. Scott, S.A., Cohen, N., Brandt, T., Warburton, P.E. and Edelmann, L. (2010) Large inverted repeats within Xp11.2 are present at the breakpoints of isodicentric X chromosomes in Turner syndrome. *Hum. Mol. Genet.*, **19**, 3383–3393.
28. 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.
29. Lieber, M.R. (2008) The mechanism of human nonhomologous DNA end joining. *J. Biol. Chem.*, **283**, 1–5.
30. Chen, J.M., Cooper, D.N., Ferec, C., Kehrer-Sawatzki, H. and Patrinos, G.P. (2010) Genomic rearrangements in inherited disease and cancer. *Semin. Cancer Biol.*, **20**, 222–233.
31. Chen, J.M., Chuzhanova, N., Stenson, P.D., Ferec, C. and Cooper, D.N. (2005) Meta-analysis of gross insertions causing human genetic disease: novel mutational mechanisms and the role of replication slippage. *Hum. Mutat.*, **25**, 207–221.
32. Chen, J.M., Chuzhanova, N., Stenson, P.D., Ferec, C. and Cooper, D.N. (2005) Complex gene rearrangements caused by serial replication slippage. *Hum. Mutat.*, **26**, 125–134.
33. Chen, J.M., Chuzhanova, N., Stenson, P.D., Ferec, C. and Cooper, D.N. (2005) Intrachromosomal serial replication slippage in trans gives rise to diverse genomic rearrangements involving inversions. *Hum. Mutat.*, **26**, 362–373.
34. Sheen, C.R., Jewell, U.R., Morris, C.M., Brennan, S.O., Ferec, C., George, P.M., Smith, M.P. and Chen, J.M. (2007) Double complex mutations involving F8 and FUNDC2 caused by distinct break-induced replication. *Hum. Mutat.*, **28**, 1198–1206.
35. 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.
36. 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.
37. 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.
38. Lobachev, K.S., Rattray, A. and Narayanan, V. (2007) Hairpin- and cruciform-mediated chromosome breakage: causes and consequences in eukaryotic cells. *Front. Biosci.*, **12**, 4208–4220.
39. Pearson, C.E., Nichol Edamura, K. and Cleary, J.D. (2005) Repeat instability: mechanisms of dynamic mutations. *Nat. Rev. Genet.*, **6**, 729–742.
40. DePamphilis, M.L. and Wassarman, P.M. (1980) Replication of eukaryotic chromosomes: a close-up of the replication fork. *Annu. Rev. Biochem.*, **49**, 627–666.
41. Ma, Y., Schwarz, K. and Lieber, M.R. (2005) The Artemis:DNA-PKcs endonuclease cleaves DNA loops, flaps, and gaps. *DNA Repair (Amst.)*, **4**, 845–851.

42. Mizuno, K., Lambert, S., Baldacci, G., Murray, J.M. and Carr, A.M. (2009) Nearby inverted repeats fuse to generate acentric and dicentric palindromic chromosomes by a replication template exchange mechanism. *Genes Dev.*, **23**, 2876–2886.
43. Paek, A.L., Kaochar, S., Jones, H., Elezaby, A., Shanks, L. and Weinert, T. (2009) Fusion of nearby inverted repeats by a replication-based mechanism leads to formation of dicentric and acentric chromosomes that cause genome instability in budding yeast. *Genes Dev.*, **23**, 2861–2875.
44. Vanneste, E., Voet, T., Le Caignec, C., Ampe, M., Konings, P., Melotte, C., Debrock, S., Amyere, M., Vikkula, M., Schuit, F. *et al.* (2009) Chromosome instability is common in human cleavage-stage embryos. *Nat. Med.*, **15**, 577–583.
45. Zhang, F., Seeman, P., Liu, P., Weterman, M.A., Gonzaga-Jauregui, C., Towne, C.F., Batish, S.D., De Vriendt, E., De Jonghe, P., Rautenstrauss, B. *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.
46. Carvalho, C.M., Zhang, F., Liu, P., Patel, A., Sahoo, T., Bacino, C.A., Shaw, C., Peacock, S., Pursley, A., Tavyev, Y.J. *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.
47. Vissers, L.E., Bhatt, S.S., Janssen, I.M., Xia, Z., Lalani, S.R., Pfundt, R., Derwinska, K., de Vries, B.B., Gilissen, C., Hoischen, A. *et al.* (2009) Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genomic architecture. *Hum. Mol. Genet.*, **18**, 3579–3593.
48. Ropers, H.H., Hoeltzenbein, M., Kalscheuer, V., Yntema, H., Hamel, B., Fryns, J.P., Chelly, J., Partington, M., Gecz, J. and Moraine, C. (2003) Nonsyndromic X-linked mental retardation: where are the missing mutations? *Trends Genet.*, **19**, 316–320.
49. Froyen, G., Corbett, M., Vandewalle, J., Jarvela, I., Lawrence, O., Meldrum, C., Bauters, M., Govaerts, K., Vandeleur, L., Van Esch, H. *et al.* (2008) Submicroscopic duplications of the hydroxysteroid dehydrogenase HSD17B10 and the E3 ubiquitin ligase HUWE1 are associated with mental retardation. *Am. J. Hum. Genet.*, **82**, 432–443.
50. Giorda, R., Bonaglia, M.C., Beri, S., Fichera, M., Novara, F., Magini, P., Urquhart, J., Sharkey, F.H., Zucca, C., Grasso, R. *et al.* (2009) Complex segmental duplications mediate a recurrent dup(X)(p11.22–p11.23) associated with mental retardation, speech delay, and EEG anomalies in males and females. *Am. J. Hum. Genet.*, **85**, 394–400.
51. Gorski, J.L., Burright, E.N., Harnden, C.E., Stein, C.K., Glover, T.W. and Reyner, E.L. (1991) Localization of DNA sequences to a region within Xp11.21 between incontinentia pigmenti (IP1) X-chromosomal translocation breakpoints. *Am. J. Hum. Genet.*, **48**, 53–64.
52. Shchelochkov, O.A., Cooper, M.L., Ou, Z., Peacock, S., Yatsenko, S.A., Brown, C.W., Fang, P., Stankiewicz, P. and Cheung, S.W. (2008) Mosaicism for r(X) and der(X)del(X)(p11.23)dup(X)(p11.21p11.22) provides insight into the possible mechanism of rearrangement. *Mol. Cytogenet.*, **1**, 16.
53. Eeles, R.A., Kote-Jarai, Z., Giles, G.G., Olama, A.A., Guy, M., Jugurnauth, S.K., Mulholland, S., Leongamornlert, D.A., Edwards, S.M., Morrison, J. *et al.* (2008) Multiple newly identified loci associated with prostate cancer susceptibility. *Nat. Genet.*, **40**, 316–321.
54. Clark, J., Rocques, P.J., Crew, A.J., Gill, S., Shipley, J., Chan, A.M., Gusterson, B.A. and Cooper, C.S. (1994) Identification of novel genes, SYT and SSX, involved in the t(X;18)(p11.2;q11.2) translocation found in human synovial sarcoma. *Nat. Genet.*, **7**, 502–508.
55. Kidd, J.M., Cooper, G.M., Donahue, W.F., Hayden, H.S., Sampas, N., Graves, T., Hansen, N., Teague, B., Alkan, C., Antonacci, F. *et al.* (2008) Mapping and sequencing of structural variation from eight human genomes. *Nature*, **453**, 56–64.
56. Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J. and Stahl, F.W. (1983) The double-strand-break repair model for recombination. *Cell*, **33**, 25–35.
57. Mann, K., Donaghue, C., Fox, S.P., Docherty, Z. and Ogilvie, C.M. (2004) Strategies for the rapid prenatal diagnosis of chromosome aneuploidy. *Eur. J. Hum. Genet.*, **12**, 907–915.
58. Jurka, J. (2000) Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet.*, **16**, 418–420.
59. Rozen, S. and Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.*, **132**, 365–386.
60. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
61. Kent, W.J. (2002) BLAT–the BLAST-like alignment tool. *Genome Res.*, **12**, 656–664.
62. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.