Genome **Medicine**

REVIEW

Inversion variants in the human genome: role in disease and genome architecture

Lars Feuk*

Abstract

Significant advances have been made over the past 5 years in mapping and characterizing structural variation in the human genome. Despite this progress, our understanding of inversion variants is still very restricted. While unbalanced variants such as copy number variations can be mapped using array-based approaches, strategies for characterization of inversion variants have been limited and underdeveloped. Traditional cytogenetic approaches have long been able to identify microscopic inversion events, but discovery of submicroscopic events has remained elusive and largely ignored. With the advent of paired-end sequencing approaches, it is now possible to map inversions across the human genome. Based on the paired-end sequencing studies published to date, it is now feasible to make a first map of inversions across the human genome and to use this map to explore the characteristics and distribution of this form of variation. The current map of inversions indicates that many remain to be identified, especially in the smaller size ranges. This review provides an overview of the current knowledge about human inversions and their contribution to human phenotypes. Further characterization of inversions should be considered as an important step towards a deeper understanding of human variation and genome dynamics.

Inversions

Over the past 5 years there has been a major drive in genomic research to identify submicroscopic structural variation in the human genome, ranging from a few hundred base pairs to approximately five megabases (Mb) in size. Structural variation is a term describing all forms of rearrangements, including deletions, duplications,

insertions, inversions, translocations and more complex rearrangements. The main type of submicroscopic variation is copy number variation (CNV) [1,2], a term used to describe gains and losses of segments of DNA. The initial reports on CNVs as an abundant form of variation in the human genome were published in 2004 [3,4]. Since then, there have been multiple studies performed to characterize the extent and importance of CNV in the human genome [5-14]. The majority of these studies have been based on microarrays, either as comparative genomic hybridization (CGH) arrays or single nucleotide polymorphism (SNP) arrays. Using array-based strategies, it is possible to identify unbalanced changes, that is, net gain or loss of large segments of DNA. However, other forms of variation involving a change in orientation or relocation of DNA, without any gain or loss, cannot readily be detected with arrays. Therefore, despite the great success in developing human genome maps of deletions and duplications, the mapping of inversions has lagged behind.

It is still not clear how many common inversions exist in the human genome, what the size distribution of inversions variants is, and to what extent inversions are associated with human disorders. With the recent introduction of novel high-throughput sequencing techniques, the methodology is now available to screen for inversions in an unbiased manner. As a consequence, our understanding of the extent of inversion variants in the human genome has increased dramatically in the past few years. This review will give an overview of the current knowledge of inversions in the human genome, the methods used to discover and type inversions, and their role in human disease and human genome architecture.

Cytogenetically visible inversions

It has long been possible to detect inversions of large chromosomal regions in G-banded karyotypes. However, this strategy is limited to identification of variants that are several megabases in size, and even significantly larger inversions may escape detection if the inverted segment leads to little difference in the banding pattern. The long history of chromosomal studies in cytogenetics has led to the identification of several inversion variants,

*Correspondence: lars.feuk@genpat.uu.se
Address: Department of Genetics and Pathology, Rudbeck Laboratory, Uppsala University, 751 85 Uppsala, Sweden

or heteromorphisms, that exist in the population but that have no clinical significance [15]. Inversions are the most common human constitutional karyotype aberration detected in cytogenetic laboratories [16]. Pericentric inversions are most frequent, often reported for chromosomes 1, 2, 3, 5, 9, 10 and 16. These are some of the most common cytogenetically visible rearrangements in humans - for example, the pericentric inversion of chromosome 9 is found in over 1% of karyotypes [17]. However, the chromosome 9 variant and many other commonly identified hetermorphisms involve only heterochromatic DNA.

The most frequently observed variant that includes euchromatic sequence is the inv(2)(p11q13), which is considered to be of no clinical significance [18]. Other events are rarer, but still frequent enough to be seen regularly in cytogenetic screening, especially in specific population groups. In addition to these common variants, numerous rare and unique inversions have been observed in individuals with no apparent phenotype. An illustrative example is inv(10)(q11.22q21.1), a 12 Mb inversion with a carrier frequency of 0.11% in the Swedish population, but with no consistent phenotype [19]. Breakpoint and haplotype analysis indicated that this is a rare variant in the population, originating from a single founder event. Due to the balanced nature of inversions, they are often of no clinical significance unless the breakpoint disrupts a gene or falls between a gene and its transcription regulatory elements. Excluding the well-established cytogenetically characterized variants, the rate of cytogenetically visible inversions reported is significantly lower than that of translocations. However, the exact rate of inversion formation is not known. A bias is likely in ascertainment of inversions in comparison to translocations, as balanced translocations lead to more reduced fitness by increased risk for an unbalanced transmission to the offspring than inversions do. Balanced translocations are therefore commonly detected as part of investigations of reproductive difficulties, while inversions with no phenotypic effect may be transmitted through many generations and never be detected, as there may be no reason for cytogenetic screening.

One of the aspects that make inversions interesting as genomic rearrangements is their role in recent primate evolution. Comparison of the human and chimpanzee genomes shows that there are nine cytogenetically visible pericentric inversions [20] and many submicroscopic inverted sequences [21]. The majority of the nine visible inversions occurred along the chimpanzee lineage, but inversions on chromosomes 1 and 18 are specific to the human lineage. These findings indicate that inversions are a type of rearrangement that occurs quite frequently in primate chromosomal evolution. Identification of a l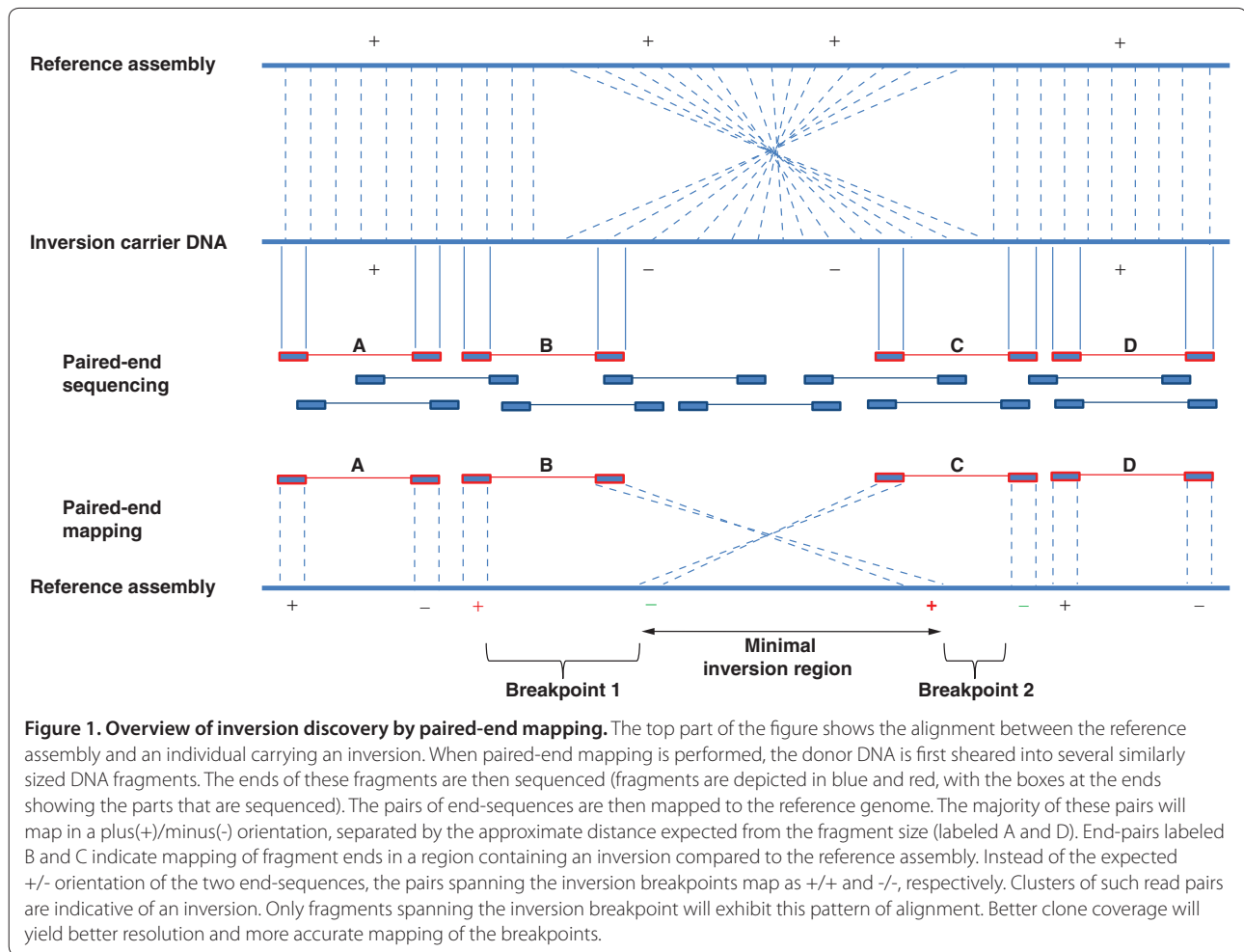arge number of inversions between closely related species, and signatures of selection associated with these, has led to speculation that inversions have played an important role in speciation [22].

Methods for inversion discovery and genotyping

Although inversions have long been detectable at the resolution of cytogenetics, progress in mapping inversions at the submicroscopic level is much more recent. As inversions only lead to a change in orientation, but not in copy number, they cannot be detected using hybridization-based methods such as microarrays. Since most strategies to map structural variation in the human genome to date have been based on array approaches, there is comparatively little known about the distribution of inversions.

Although there has been a lack of methods for global discovery of inversions, it has long been possible to test for the presence of inversions in a targeted manner if there is a prior hypothesis that a region may be inverted. Testing can be done using traditional molecular approaches such as pulse-field gel electrophoresis (PFGE) or Southern blot. Single molecular haplotyping has also been successfully used to screen samples for specific inversion variants [23]. However, these strategies are laborious and do not work for global unbiased discovery of new inversion regions on a genome-wide scale. Despite these limitations, a small number of studies have led to the identification of inversion variants using 'genomic' strategies. One approach that led to the identification of three polymorphic inversions was based on investigating regions that are inverted between the human and chimpanzee genomes. By targeting 23 such regions in human control samples, three inversions were found to be polymorphic in humans. In another study, Bansal *et al.* [24] used the linkage disequilibrium (LD) pattern of SNPs to map putative inversion breakpoints. By using a statistical method to detect regions where SNPs at a distance from each other on the reference assembly were in higher LD than SNPs in close proximity, a number of putative inversions were identified. Overlap with several previously validated inversions indicated that the approach was successful. However, the candidate variants identified by this method require experimental validation to distinguish real inversions from false positives. Although the approaches outlined above have shown some success in the discovery of novel inversion variants, recent data indicate that only a very small fraction of frequent human inversions were found.

A major breakthrough in the discovery of inversions (and other forms of structural variation) came with the introduction of paired-end sequencing and mapping [7]. Generally, when the two ends of a cloned fragment are sequenced, the two resulting sequences would be expected to align to the reference genome in a + and - orientation,

Figure 1. Overview of inversion discovery by paired-end mapping. The top part of the figure shows the alignment between the reference assembly and an individual carrying an inversion. When paired-end mapping is performed, the donor DNA is first sheared into several similarly sized DNA fragments. The ends of these fragments are then sequenced (fragments are depicted in blue and red, with the boxes at the ends showing the parts that are sequenced). The pairs of end-sequences are then mapped to the reference genome. The majority of these pairs will map in a plus(+)/minus(-) orientation, separated by the approximate distance expected from the fragment size (labeled A and D). End-pairs labeled B and C indicate mapping of fragment ends in a region containing an inversion compared to the reference assembly. Instead of the expected +/- orientation of the two end-sequences, the pairs spanning the inversion breakpoints map as +/+ and -/-, respectively. Clusters of such read pairs are indicative of an inversion. Only fragments spanning the inversion breakpoint will exhibit this pattern of alignment. Better clone coverage will yield better resolution and more accurate mapping of the breakpoints.

respectively. However, if the donor DNA carries an inversion as compared to the reference assembly, this would lead to the end sequences of fragments spanning the breakpoints to align in a -/- or a +/+ orientation (Figure 1). By searching for clusters of fragments exhibiting this pattern of alignments to the reference assembly, it is possible to identify putative inversion events. The first paired-end mapping study was based on end sequencing of fosmid clones using traditional Sanger sequencing [7]. The study identified 56 inversion break-points from a fosmid library representing a single human genome (sample NA15510). The same strategy of fosmid end sequencing was later applied to another eight genomes, and a total of 217 inversions were identified and validated [6]. A large number of inversions were also reported in the first individual genome to be sequenced (the genome of Craig Venter, called HuRef) [25]. Sanger sequencing was employed to sequence the HuRef genome, and an assembly was created independently from the National Center for Biotechnology Information (NCBI) reference assembly. An assembly comparison

analysis gave rise to 90 regions of inverted orientation between the HuRef and NCBI assemblies. Since these initial Sanger sequencing studies, the general strategy of paired-end mapping has been adapted to fragment end-sequencing with second-generation-sequencing platforms [26,27]. Although only a small number of whole-genome sequencing studies have so far employed this strategy to identify inversions, this is likely to be the main approach for identification of inversions in the near future.

Despite the success of paired-end mapping, there are still challenges to overcome. One important feature of the paired-end mapping approach is that it relies on the reference assembly. It is well established that the reference assembly represents very rare or unique alleles at some loci in the genome. In rare instances, it is also possible that these unique alleles represent cloning artifacts or are a result of mis-assembly of the reference sequence. For example, this has been suggested for an inversion overlapping an exon of the *DOCK3* gene on chromosome 3, for which there is an inversion in the reference assembly as compared to available mRNA

sequences for the same gene [5]. For regions where the reference assembly harbors a unique allele, every study with high enough resolution and sequence coverage will identify a homozygous inversion.

Another limitation of paired-end mapping for inversion detection is related to the genome architecture associated with inversions. The majority of large (>100 kb) inversions described in the human genome to date are flanked by high identity segmental duplications, that is, sequences >1 kb that exist in two or more copies of >90% identity in the human genome [28,29]. The segmental duplications associated with inversions cause problems for inversion discovery using paired-end mapping. As the method depends on alignment to the reference assembly, highly identical sequences in the assembly will cause problems in identifying unique placements for the sequence reads. Many paired-end mapping pipelines simply discard reads that cannot be uniquely mapped. Therefore, the paired-end mapping strategy often fails to identify inversions flanked by long inverted segmental duplications of high identity. For these regions, targeted assays are required.

Current map of inversions in the human genome

The map of human inversions is still quite limited, and our understanding of the number of inversions, the size distribution and the frequency distribution is probably biased due to biases in the approaches used for variation identification. There are currently 914 inversion events reported in the Database of Genomic Variants [30], a database resource for structural variation in the human genome [3,31]. However, many of these overlap and actually refer to the same locus. If only non-redundant loci are counted, there are a total of 479 inversions in the database. Figure 2 shows an overview of the current inversions reported in the human genome. The inversions are found across the size spectrum up to several megabases. A comparison of the size distribution of inversions and CNVs is shown in Figure 3. The size distribution shows that most of the inversions discovered to date are in the 10 kb to 100 kb interval. For CNVs, size distribution is shifted more towards smaller size variants.

There are many potential explanations for the difference in size distribution between inversions and CNVs (Figure 3). Biologically, large inversions are more likely to be neutral, without obvious phenotypic consequences, compared to large CNVs. Data from cytogenetic studies support this. One difference between inversions and CNVs is that the genes within an inversion can be entirely unaffected, while genes within CNVs are always affected by a dosage imbalance. For inversions, it is more important where the breakpoints are located and if these interrupt a gene or lead to disruption of the transcriptional regulation of genes. If no gene or regulatory
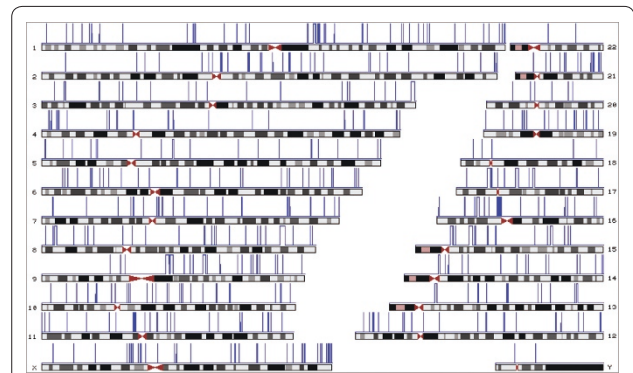
Figure 2. Distribution of inversion variants in the human genome. The blue lines in this ideogram show the human chromosomal distribution of the 479 non-redundant inversion variants reported in the Database of Genomic Variants.
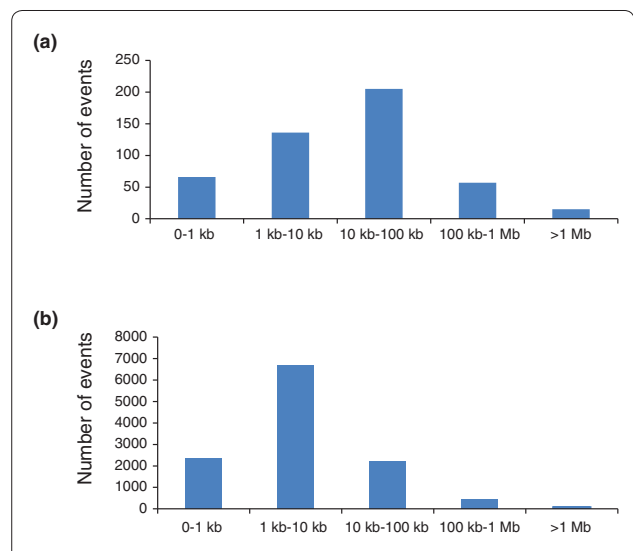
Figure 3. Size distribution of inversions and copy number variants. The size distribution of inversions reported in the Database of Genomic Variants **(a)** shows that the majority of inversions reported to date are in the 10 to 100 kb size bin. The size distribution of inversions differs from that reported for copy number variants (CNVs) **(b)** The CNV data plotted here show the 11,700 non-redundant CNV events reported by Conrad *et al.* [13]. It is currently unclear whether the difference in size distribution between inversions and CNVs is due to ascertainment bias, or whether there is an actual biological difference in size distribution. Both cytogenetic data and evolutionary comparative genomic data indicate that large inversions are less detrimental than large deletions and duplications.

function is interrupted by the breakpoints, inversions that are comparatively large may be frequent in the population. While there are very few CNVs >1 Mb in size that have reached a minor allele frequency of 1%, there are examples of very large inversions that are frequently observed in the population. The best-studied examples are two inversions located on chromosomes 4 and 8, respectively. Both these inversions have breakpoints that

fall in clusters of olfactory receptors of high identity. The inversion on chromosome 8 is approximately 3.5 Mb in size and has been reported to be present in 26% of healthy controls, while the chromosome 4 inversion is about 6 Mb in size and was found in 12.5% of healthy controls [32]. These data indicate that very large inversions may exist in the human genomes without a strong negative effect on reproductive fitness.

There may also be a methodological explanation for the difference in size distribution between current annotations of inversions and CNVs, based on differences in methods of discovery and limitations in technology. The size distribution for inversions is reflective of the resolution and limited sequence coverage of the paired-end mapping projects published to date. For very small inversions, deep sequence coverage would be required to obtain several DNA fragments spanning one breakpoint. Therefore, many additional inversions will be found as thousands of additional genomes are sequenced over the next few years, and a large fraction of these would be expected to increase the fraction of variants that are <10 kb in size.

Finally, it is also possible that the size distribution for inversions differs from that of CNVs based on the mechanisms by which the variants are created. As for CNVs [13], it is likely that different mechanisms act across the size spectrum and give rise to larger and smaller inversion events, respectively. Through non-allelic homologous recombination (NAHR) - recombination events taking place between highly similar sequences - regions located between segmental duplications or highly identical repeat sequences may be deleted, duplicated or inverted. Inversions can be formed by this process if the duplicated sequences are in inverted orientation with respect to each other. Therefore, NAHR is considered the primary mechanism by which large (tens of kilobases) inversions are formed. However, for small inversions, the mechanisms are not as well characterized as for smaller insertions/deletions. Some evidence points towards replication-based mechanisms, such as microhomology-mediated break-induced replication (MMBIR) [33]. Other specific mechanisms that have been suggested to be involved in creation of inversions include fork stalling and template switching (FoSTeS) [34] and serial replication slippage *in trans* [35]. However, the limited number of inversions with nucleotide resolution breakpoint information available to date has prevented a thorough investigation of mechanisms and sequence motifs giving rise to inversions. As additional inversion breakpoints are identified, these relationships should become more evident.

Inversions in human disorders

There are many descriptions in the literature of patients with specific phenotypes who also carry an inversion that is cytogenetically visible. Since inversions are relatively rare events, and it is unlikely that multiple patients with the same inversion are found, it is often problematic to assess whether the inversion present in the patient is actually associated with the phenotype. The exception is if the inversion breakpoint falls within or near a gene that has previously been associated with the disorder through other types of mutations. For recurrent inversions, the association between phenotype and genotype is more obvious, and a number of such loci have been described. One of the best-characterized recurrent inversions giving rise to disease causes hemophilia A, an X-linked disorder caused by mutations in the factor VIII gene [36]. A recurrent inversion has been found in approximately 43% of patients [37]. Molecular characterization of the breakpoints indicates that the inversion is a result of intrachromosomal homologous recombination, originating almost exclusively in male germ cells. This recurrent inversion spans approximately 400 kb and is mediated by two inverted segmental duplications, one of which is located in intron 22 of the factor VIII gene, with two other copies being located approximately 400 kb telomeric to the gene. Other examples where recurrent inversions have been shown to lead to a disease phenotype are the disruption of the idunorate 2-sulphatase gene in mucopolysaccharidosis type II (Hunter syndrome) [38], and disruption of the emerin gene in Emery-Dreifuss muscular dystrophy [39].

A specific category of inversions associated with genetic disorders is those that are not directly causative, but rather increase the risk of further rearrangements that cause disease. For a number of microdeletion syndromes, one or both parents of probands have been found to carry an inversion of the deleted interval. The association was first described in Williams-Beuren syndrome, which is most commonly caused by a 1.5 Mb microdeletion at 7q11. In a study of 12 families where the proband carried the typical microdeletion, an inversion was found in a parent for 33% of the patients [40]. The inversion variant has since been shown to be relatively frequent in the general population (approximately 5%), and does not seem to be associated with a phenotype in itself [41].

Another example of a disorder where an inversion has been associated with a causative deletion is the 17q21.31 microdeletion syndrome, a genetically characterized form of mental retardation. This region harbors a 970 kb inversion polymorphism found at high frequency in European populations [42]. The genetic variation pattern within the region indicates that the inversion first appeared before dispersal out of Africa, and that there has been little or no recombination between the haplotypes. Interestingly, there is some evidence that this inversion variation is associated with higher reproductive

Table 1. Rearrangements associated with inversion variants

Chromosome band	Inversion size (Mb)	Disorder/rearrangement	Reference (syndrome : inversion)
3q29	1.9	3q29 deletion syndrome	[49] : [7]
5q35.2-q35.3*	1.9	Sotos syndrome microdeletion	[50] : [51]
7q11.23*	1.5	Williams-Beuren syndrome microdeletion	[52] : [40]
8p23[a]	4.7	Inv dup(8p) and del (8)(p23.1;p23.2)	[53,54] : [32,55]
15q11-q13*	4	Angelman syndrome deletion	[56] : [57]
15q13.3*	2	15q13.3 microdeletion	[58] : [6,58]
15q24	1.2	15q24 microdeletion	[44,59] : [6]
17q12	1.5	Renal cysts and diabetes (RCAD) microdeletion syndrome	[60] : [6]
17q21.31*	0.9	17q21.31 microdeletion syndrome	[43-45] : [42]

[a]The inversion has been found at higher frequency in parents of probands with microdeletions than in the general population, indicating that the inversion is a risk factor for subsequent rearrangements in the offspring.

fitness [42]. Screening patient cohorts with mental retardation led to the discovery of a microdeletion syndrome corresponding to the same region as the common inversion polymorphism [43-45]. Studies of the parents of microdeletion carriers showed that at least one parent carried the inverted H2 haplotype in every case. It was therefore initially concluded that the inversion in itself was the cause of the increased risk for the deletion to occur. It has been suggested that the lack of homology across the inversion region between heterozygous chromatids in meiosis may lead to the formation of an 'asynaptic bubble' that renders the region unstable and prone to additional rearrangements [46]. However, additional characterization of the prevalent haplotypes in the region indicates that other rearrangements present on the inverted H2 haplotype may be the primary substrate for the non-allelic homologous recombination giving rise to the microdeletion [47]. Additional studies will be needed to confirm exactly how the inversion leads to an increased risk for deletions in the offspring.

In total, there are at least nine different microdeletion syndromes for which the deletion region has also been found as an inversion variant in the general population (Table 1). For a majority of these disorders, a direct association between the inversion carrier status and increased risk for deletion in the offspring has been established by comparing the inversion frequency in parents to the frequency in the general population. However, the exact molecular mechanisms still remain to be elucidated and it is not confirmed whether it is the inversion itself, or other sequence features present on the inversion haplotype, that causes the subsequent pathogenic rearrangement.

Conclusions and future perspectives

With the advent of deep coverage paired-end sequencing, the number of inversions reported has increased dramatically and the inversion breakpoints will be pinpointed at much higher resolution. Over the next year or two, the true extent of inversion variants in the human genome will be revealed. Only then will it be possible to explore the contribution of inversions to common disease. For both inversions and other structural variants, it has been anticipated that it would be possible to impute these variants from high-density SNP array data. However, recent studies indicate that this may not be the case. Data from one study show that many large inversions, surrounded by blocks of segmental duplications, have arisen on more than one haplotype background [48]. Similar data have been shown for multi-allelic CNVs [13]. These variants will therefore need to be directly targeted for inclusion in association studies. Currently, the experimental strategies for accurate high-throughput genotyping of inversions and multi-allelic CNVs are limited or non-existent. However, it is very likely that smaller inversions that are not flanked by blocks of segmental duplications will have arisen only once and will therefore be in LD with surrounding SNPs. This has been shown in a limited number of cases [21], but more data are needed to confirm whether this applies to a majority of events. Other questions that remain to be explored in further detail include inversion formation mechanisms, characterization of breakpoints, and development of maps and strategies for inclusion of inversion variants in genome-wide disease association studies. In conclusion, we are now at the stage where we have the tools that enable characterization of the full extent of inversions in the human genome and their contribution to human variation and disease.

Abbreviations
CGH, comparative genomic hybridization; CNV, copy number variation; FoSTeS, fork stalling and template switching; kb, kilobase; LD, linkage disequilibrium; Mb, megabase; MMBIR, microhomology-mediated break-induced replication; NAHR, non-allelic homologous recombination; NCBI, National Center for Biotechnology Information; PFGE, pulse-field gel electrophoresis; SNP, single nucleotide polymorphism.

References
1. Feuk L, Carson AR, Scherer SW: **Structural variation in the human genome.** *Nat Rev Genet* 2006, **7:**85-97.
2. Sharp AJ, Cheng Z, Eichler EE: **Structural variation of the human genome.** *Annu Rev Genomics Hum Genet* 2006, **7:**407-442.
3. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C: **Detection of large-scale variation in the human genome.** *Nat Genet* 2004, **36:**949-951.
4. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M: **Large-scale copy number polymorphism in the human genome.** *Science* 2004, **305:**525-528.
5. Khaja R, Zhang J, MacDonald JR, He Y, Joseph-George AM, Wei J, Rafiq MA, Qian C, Shago M, Pantano L, Aburatani H, Jones K, Redon R, Hurles M, Armengol L, Estivill X, Mural RJ, Lee C, Scherer SW, Feuk L: **Genome assembly comparison identifies structural variants in the human genome.** *Nat Genet* 2006, **38:**1413-1418.
6. Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, Hansen N, Teague B, Alkan C, Antonacci F, Haugen E, Zerr T, Yamada NA, Tsang P, Newman TL, Tuzun E, Cheng Z, Ebling HM, Tusneem N, David R, Gillett W, Phelps KA, Weaver M, Saranga D, Brand A, Tao W, Gustafson E, McKernan K, Chen L, Malig M, *et al.*: **Mapping and sequencing of structural variation from eight human genomes.** *Nature* 2008, **453:**56-64.
7. Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, Eichler EE: **Fine-scale structural variation of the human genome.** *Nat Genet* 2005, **37:**727-732.
8. McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, Dallaire S, Gabriel SB, Lee C, Daly MJ, Altshuler DM: **Common deletion polymorphisms in the human genome.** *Nat Genet* 2006, **38:**86-92.
9. McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, Shapero MH, de Bakker PI, Maller JB, Kirby A, Elliott AL, Parkin M, Hubbell E, Webster T, Mei R, Veitch J, Collins PJ, Handsaker R, Lincoln S, Nizzari M, Blume J, Jones KW, Rava R, Daly MJ, Gabriel SB, Altshuler D: **Integrated detection and population-genetic analysis of SNPs and copy number variation.** *Nat Genet* 2008, **40:**1166-1174.
10. Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK: **A high-resolution survey of deletion polymorphism in the human genome.** *Nat Genet* 2006, **38:**75-81.
11. Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R, Oseroff VV, Albertson DG, Pinkel D, Eichler EE: **Segmental duplications and copy-number variation in the human genome.** *Am J Hum Genet* 2005, **77:**78-88.
12. Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR: **Whole-genome patterns of common DNA variation in three human populations.** *Science* 2005, **307:**1072-1079.
13. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, Macarthur DG, Macdonald JR, Onyiah I, Pang AW, Robson S, Stirrups K, Valsesia A, Walter K, Wei J, Tyler-Smith C, Carter NP, Lee C, Scherer SW, Hurles ME: **Origins and functional impact of copy number variation in the human genome.** *Nature* 2009 [Epub ahead of print].
14. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, *et al.*: **Global variation in copy number in the human genome.** *Nature* 2006, **444:**444-454.
15. Thomas NS, Bryant V, Maloney V, Cockwell AE, Jacobs PA: **Investigation of the origins of human autosomal inversions.** *Hum Genet* 2008, **123:**607-616.
16. Schmidt S, Claussen U, Liehr T, Weise A: **Evolution versus constitution: differences in chromosomal inversion.** *Hum Genet* 2005, **117:**213-219.
17. Hsu LY, Benn PA, Tannenbaum HL, Perlis TE, Carlson AD: **Chromosomal polymorphisms of 1, 9, 16, and Y in 4 major ethnic groups: a large prenatal study.** *Am J Med Genet* 1987, **26:**95-101.
18. MacDonald IM, Cox DM: **Inversion of chromosome 2 (p11p13): frequency and implications for genetic counselling.** *Hum Genet* 1985, **69:**281-283.
19. Entesarian M, Carlsson B, Mansouri MR, Stattin EL, Holmberg E, Golovleva I, Stefansson H, Klar J, Dahl N: **A chromosome 10 variant with a 12 Mb inversion [inv(10)(q11.22q21.1)] identical by descent and frequent in the Swedish population.** *Am J Med Genet A* 2009, **149A:**380-386.
20. Yunis JJ, Prakash O: **The origin of man: a chromosomal pictorial legacy.** *Science* 1982, **215:**1525-1530.
21. Feuk L, Macdonald JR, Tang T, Carson AR, Li M, Rao G, Khaja R, Scherer SW: **Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies.** *PLoS Genet* 2005, **1:**e56.
22. Navarro A, Barton NH: **Chromosomal speciation and molecular divergence--accelerated evolution in rearranged chromosomes.** *Science* 2003, **300:**321-324.
23. Turner DJ, Shendure J, Porreca G, Church G, Green P, Tyler-Smith C, Hurles ME: **Assaying chromosomal inversions by single-molecule haplotyping.** *Nat Methods* 2006, **3:**439-445.
24. Bansal V, Bashir A, Bafna V: **Evidence for large inversion polymorphisms in the human genome from HapMap data.** *Genome Res* 2007, **17:**219-230.
25. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers YH, Frazier ME, Scherer SW, Strausberg RL, *et al.*: **The diploid genome sequence of an individual human.** *PLoS Biol* 2007, **5:**e254.
26. Ahn SM, Kim TH, Lee S, Kim D, Ghang H, Kim DS, Kim BC, Kim SY, Kim WY, Kim C, Park D, Lee YS, Kim S, Reja R, Jho S, Kim CG, Cha JY, Kim KH, Lee B, Bhak J, Kim SJ: **The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group.** *Genome Res* 2009, **19:**1622-1629.
27. Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, Taillon BE, Chen Z, Tanzer A, Saunders AC, Chi J, Yang F, Carter NP, Hurles ME, Weissman SM, Harkins TT, Gerstein MB, Egholm M, Snyder M: **Paired-end mapping reveals extensive structural variation in the human genome.** *Science* 2007, **318:**420-426.
28. Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE: **Recent segmental duplications in the human genome.** *Science* 2002, **297:**1003-1007.
29. Eichler EE: **Segmental duplications: what's missing, misassigned, and misassembled - and should we care?** *Genome Res* 2001, **11:**653-656.
30. Database of Genomics Variants [http://projects.tcag.ca/variation/] (accessed 2 February 2010).
31. Zhang J, Feuk L, Duggan GE, Khaja R, Scherer SW: **Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome.** *Cytogenet Genome Res* 2006, **115:**205-214.
32. Giglio S, Calvari V, Gregato G, Gimelli G, Camanini S, Giorda R, Ragusa A, Guerneri S, Selicorni A, Stumm M, Tonnies H, Ventura M, Zollino M, Neri G, Barber J, Wieczorek D, Rocchi M, Zuffardi O: **Heterozygous submicroscopic inversions involving olfactory receptor-gene clusters mediate the recurrent t(4;8)(p16;p23) translocation.** *Am J Hum Genet* 2002, **71:**276-285.
33. Hastings PJ, Ira G, Lupski JR: **A microhomology-mediated break-induced replication model for the origin of human copy number variation.** *PLoS Genet* 2009, **5:**e1000327.
34. Zhang F, Khajavi M, Connolly AM, Towne CF, Batish SD, Lupski JR: **The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans.** *Nat Genet* 2009, **41:**849-853.
35. Chen JM, Chuzhanova N, Stenson PD, Ferec C, Cooper DN: **Intrachromosomal serial replication slippage in trans gives rise to diverse genomic rearrangements involving inversions.** *Hum Mutat* 2005, **26:**362-373.
36. Lakich D, Kazazian HH Jr, Antonarakis SE, Gitschier J: **Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A.** *Nat Genet* 1993, **5:**236-241.
37. Antonarakis SE, Rossiter JP, Young M, Horst J, de Moerloose P, Sommer SS, Ketterling RP, Kazazian HH Jr, Negrier C, Vinciguerra C, Gitschier J, Goossens M, Girodon E, Ghanem N, Plassa F, Lavergne JM, Vidaud M, Costa JM, Laurian Y, Lin SW, Lin SR, Shen MC, Lillicrap D, Taylor SA, Windsor S, Valleix SV, Nafa K, Sultan Y, Delpech M, Vnencak-Jones CL, *et al.*: **Factor VIII gene inversions in**

severe hemophilia A: results of an international consortium study. *Blood* 1995, **86**:2206-2212.

38. Bondeson ML, Dahl N, Malmgren H, Kleijer WJ, Tonnesen T, Carlberg BM, Pettersson U: Inversion of the IDS gene resulting from recombination with IDS-related sequences is a common cause of the Hunter syndrome. *Hum Mol Genet* 1995, **4**:615-621.

39. Small K, Iber J, Warren ST: Emerin deletion reveals a common X-chromosome inversion mediated by inverted repeats. *Nat Genet* 1997, **16**:96-99.

40. Osborne LR, Li M, Pober B, Chitayat D, Bodurtha J, Mandel A, Costa T, Grebe T, Cox S, Tsui LC, Scherer SW: A 1.5 million-base pair inversion polymorphism in families with Williams-Beuren syndrome. *Nat Genet* 2001, **29**:321-325.

41. Tam E, Young EJ, Morris CA, Marshall CR, Loo W, Scherer SW, Mervis CB, Osborne LR: The common inversion of the Williams-Beuren syndrome region at 7q11.23 does not cause clinical symptoms. *Am J Med Genet A* 2008, **146A**:1797-1806.

42. Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, Baker A, Jonasdottir A, Ingason A, Gudnadottir VG, Desnica N, Hicks A, Gylfason A, Gudbjartsson DF, Jonsdottir GM, Sainz J, Agnarsson K, Birgisdottir B, Ghosh S, Olafsdottir A, Cazier JB, Kristjansson K, Frigge ML, Thorgeirsson TE, Gulcher JR, Kong A, Stefansson K: A common inversion under selection in Europeans. *Nat Genet* 2005, **37**:129-137.

43. Koolen DA, Vissers LE, Pfundt R, de Leeuw N, Knight SJ, Regan R, Kooy RF, Reyniers E, Romano C, Fichera M, Schinzel A, Baumer A, Anderlid BM, Schoumans J, Knoers NV, van Kessel AG, Sistermans EA, Veltman JA, Brunner HG, de Vries BB: A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. *Nat Genet* 2006, **38**:999-1001.

44. Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, Hurst JA, Stewart H, Price SM, Blair E, Hennekam RC, Fitzpatrick CA, Segraves R, Richmond TA, Guiver C, Albertson DG, Pinkel D, Eis PS, Schwartz S, Knight SJ, Eichler EE: Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat Genet* 2006, **38**:1038-1042.

45. Shaw-Smith C, Pittman AM, Willatt L, Martin H, Rickman L, Gribble S, Curley R, Cumming S, Dunn C, Kalaitzopoulos D, Porter K, Prigmore E, Krepischi-Santos AC, Varela MC, Koiffmann CP, Lees AJ, Rosenberg C, Firth HV, de Silva R, Carter NP: Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. *Nat Genet* 2006, **38**:1032-1037.

46. Sharp AJ: Emerging themes and new challenges in defining the role of structural variation in human disease. *Hum Mutat* 2009, **30**:135-144.

47. Zody MC, Jiang Z, Fung HC, Antonacci F, Hillier LW, Cardone MF, Graves TA, Kidd JM, Cheng Z, Abouelleil A, Chen L, Wallis J, Glasscock J, Wilson RK, Reily AD, Duckworth J, Ventura M, Hardy J, Warren WC, Eichler EE: Evolutionary toggling of the MAPT 17q21.31 inversion region. *Nat Genet* 2008, **40**:1076-1083.

48. Antonacci F, Kidd JM, Marques-Bonet T, Ventura M, Siswara P, Jiang Z, Eichler EE: Characterization of six human disease-associated inversion polymorphisms. *Hum Mol Genet* 2009, **18**:2555-2566.

49. Willatt L, Cox J, Barber J, Cabanas ED, Collins A, Donnai D, FitzPatrick DR, Maher E, Martin H, Parnau J, Pindar L, Ramsay J, Shaw-Smith C, Sistermans EA, Tettenborn M, Trump D, de Vries BB, Walker K, Raymond FL: 3q29 microdeletion syndrome: clinical and molecular characterization of a new syndrome. *Am J Hum Genet* 2005, **77**:154-160.

50. Kurotaki N, Imaizumi K, Harada N, Masuno M, Kondoh T, Nagai T, Ohashi H, Naritomi K, Tsukahara M, Makita Y, Sugimoto T, Sonoda T, Hasegawa T, Chinen Y, Tomita Ha HA, Kinoshita A, Mizuguchi T, Yoshiura Ki K, Ohta T, Kishino T, Fukushima Y, Niikawa N, Matsumoto N: Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat Genet* 2002, **30**:365-366.

51. Visser R, Shimokawa O, Harada N, Kinoshita A, Ohta T, Niikawa N, Matsumoto N: Identification of a 3.0-kb major recombination hotspot in patients with sotos syndrome who carry a common 1.9-Mb microdeletion. *Am J Hum Genet* 2005, **76**:52-67.

52. Ewart AK, Morris CA, Atkinson D, Jin W, Sternes K, Spallone P, Stock AD, Leppert M, Keating MT: Hemizygosity at the elastin locus in a developmental disorder, Williams syndrome. *Nat Genet* 1993, **5**:11-16.

53. Devriendt K, Matthijs G, Van Dael R, Gewillig M, Eyskens B, Hjalgrim H, Dolmer B, McGaughran J, Brondum-Nielsen K, Marynen P, Fryns JP, Vermeesch JR: Delineation of the critical deletion region for congenital heart defects, on chromosome 8p23.1. *Am J Hum Genet* 1999, **64**:1119-1126.

54. Floridia G, Piantanida M, Minelli A, Dellavecchia C, Bonaglia C, Rossi E, Gimelli G, Croci G, Franchi F, Gilgenkrantz S, Grammatico P, Dalpra L, Wood S, Danesino C, Zuffardi O: The same molecular mechanism at the maternal meiosis I produces mono- and dicentric 8p duplications. *Am J Hum Genet* 1996, **58**:785-796.

55. Giglio S, Broman KW, Matsumoto N, Calvari V, Gimelli G, Neumann T, Ohashi H, Voullaire L, Larizza D, Giorda R, Weber JL, Ledbetter DH, Zuffardi O: Olfactory receptor-gene clusters, genomic-inversion polymorphisms, and common chromosome rearrangements. *Am J Hum Genet* 2001, **68**:874-883.

56. Knoll JH, Nicholls RD, Magenis RE, Graham JM Jr, Lalande M, Latt SA: Angelman and Prader-Willi syndromes share a common chromosome 15 deletion but differ in parental origin of the deletion. *Am J Med Genet* 1989, **32**:285-290.

57. Gimelli G, Pujana MA, Patricelli MG, Russo S, Giardino D, Larizza L, Cheung J, Armengol L, Schinzel A, Estivill X, Zuffardi O: Genomic inversions of human chromosome 15q11-q13 in mothers of Angelman syndrome patients with class II (BP2/3) deletions. *Hum Mol Genet* 2003, **12**:849-858.

58. Sharp AJ, Mefford HC, Li K, Baker C, Skinner C, Stevenson RE, Schroer RJ, Novara F, De Gregori M, Ciccone R, Broomer A, Casuga I, Wang Y, Xiao C, Barbacioru C, Gimelli G, Bernardina BD, Torniero C, Giorda R, Regan R, Murday V, Mansour S, Fichera M, Castiglia L, Failla P, Ventura M, Jiang Z, Cooper GM, Knight SJ, Romano C, *et al.*: A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat Genet* 2008, **40**:322-328.

59. Sharp AJ, Selzer RR, Veltman JA, Gimelli S, Gimelli G, Striano P, Coppola A, Regan R, Price SM, Knoers NV, Eis PS, Brunner HG, Hennekam RC, Knight SJ, de Vries BB, Zuffardi O, Eichler EE: Characterization of a recurrent 15q24 microdeletion syndrome. *Hum Mol Genet* 2007, **16**:567-572.

60. Mefford HC, Clauin S, Sharp AJ, Moller RS, Ullmann R, Kapur R, Pinkel D, Cooper GM, Ventura M, Ropers HH, Tommerup N, Eichler EE, Bellanne-Chantelot C: Recurrent reciprocal genomic rearrangements of 17q12 are associated with renal disease, diabetes, and epilepsy. *Am J Hum Genet* 2007, **81**:1057-1069.