Mapping translocation breakpoints by next-generation sequencing

Wei Chen, Vera Kalscheu, Andreas Tzschach, Corinna Menzel, Reinhard Ullmann, Marcel Schulz, Fikret Erdogan, Na Li, Zofia Kijas, Ger Arkesteijn, Isidora Lopez Pajares, Margret Goetz-Sothmann, Uwe Heinrich, Imma Rost, Andreas Dufke, Ute Grasshoff, Birgitta Glaeser Glaeser, Martin Vingron and Hans Hilger Ropers

Genome Res. published online Mar 7, 2008;
Access the most recent version at doi:10.1101/gr.076166.108

Mapping translocation breakpoints by next-generation sequencing

Wei Chen[1], Vera Kalscheuer[1], Andreas Tzschach[1], Corinna Menzel[1], Reinhard Ullmann[1], Marcel Holger Schulz[1,7], Fikret Erdogan[1], Na Li[1], Zofia Kijas[1], Ger Arkesteijn[2], Isidora Lopez Pajares[3] ,Margret Goetz-Sothmann[4],Uwe Heinrich[5] ,Imma Rost[5] , Andreas Dufke[6], Ute Grasshoff[6], Birgitta Glaeser[8], Martin Vingron[1], H. Hilger Ropers[1]

[1]Max Planck Institute for Molecular Genetics, Berlin, Germany

[2]Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands

[3]Department of Genetics, Hospital Universitario La Paz, Madrid, Spain

[4]Practice of Human Genetics, Munich, Germany

[5]Centre of Human Genetics, Martinsried, Germany

[6] Department of Medical Genetics, Eberhard Karls University, Tuebingen, Germany

[7] International Max-Planck Research School for Computational Biology and Scientific Computing

[8]Institut für Klinische Genetik, Olgahospital ,Stuttgart, Germany

[§]Correspondence: Dr. Wei Chen, Max Planck Institute for Molecular Genetics, Ihnestrasse 73, 14195 Berlin, Germany, Tel: +49 30 84131214; Fax: +49 30 84131383, E-mail: wei@molgen.mpg.de

Running title

Translocation breakpoint mapping

Keyword

Breakpoint mapping, next-generation sequencing technology

Abstract

Balanced chromosome rearrangements (BCRs) can cause genetic diseases by disrupting or inactivating specific genes, and the characterisation of breakpoints in disease-associated BCRs has been instrumental in the molecular elucidation of a wide variety of genetic disorders. However, mapping chromosome breakpoints using traditional methods, such as *in situ* hybridisation with fluorescent dye-labeled bacterial artificial chromosome clones (BAC-FISH), is rather laborious and time consuming. In addition, the resolution of BAC-FISH is often insufficient to unequivocally identify the disrupted gene. To overcome these limitations, we have performed shotgun sequencing of flow-sorted derivative chromosomes using 'next generation' (Solexa/Illumina) multiplex sequencing-by-synthesis technology. As shown here for three different disease-associated BCRs, the coverage attained by this platform is sufficient to bridge the breakpoints by PCR amplification, and this procedure allows to determine their exact nucleotide positions within few weeks. Its implementation will greatly facilitate large-scale breakpoint mapping and gene finding in patients with disease-associated balanced translocations.

Solexa sequencing data have been submitted to the Short Read Archive at NCBI http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?) and are accessible through accession number SRA00261. ArrayCGH data have been submitted to the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession number GSE10115. A detailed description of the statistical methods used is provided online as Supplementary Information. A supplementary table lists all primer pairs used to amplify the junction fragments. Flow sorting diagrams are provided as supplementary figures.

.

Introduction

De novo balanced translocations have a prevalence of approximately 1/2000 live births, and about 6% of these cases are associated with congenital malformations and/or developmental delay (Warburton 1991). This proportion is about twice as high as in the general population, suggesting a causative link between the rearrangements and the observed phenotype in at least half of the disease-associated balanced translocations (DBCRs). In most of the patients with DBCRs, aberrant phenotypes are due to truncation of genes, and 50% of these patients with DBCRs are mentally retarded (Bugge et al. 2000). Therefore, DBCRs are a unique resource for bridging genotypes and phenotypes in patients with mental retardation and other early-onset Mendelian disorders, but DBCRs have also been characterized to search for genetic risk factors for complex and late-onset diseases, as discussed below.

Mapping chromosome breakpoints using traditional methods, such as FISH, is rather laborious and time consuming; moreover, the resolution of this method is often limited, particularly if fluorescence-labelled BAC probes are employed. With the development of the array painting technique, the efficiency has been greatly improved. Upon array painting, the two derivative chromosomes are isolated by flow sorting, differentially labelled and hybridised onto DNA arrays. Since the chromosome segments located on either side of breakpoint are labelled with different fluorescent dyes, plotting fluorescence ratio of probes ordered along the chromosome can reveal the breakpoint flanking region, while breakpoint spanning probes usually show intermediate ratios (Backx et al. 2007; Fiegler et al. 2003; Veltman et al. 2003). Only recently, the resolution of this method has been increased by sequential painting with two arrays, a tiling path large insert array and a region-specific, ultra-high-resolution oligonucleotide array (Gribble et al. 2007).

At least theoretically, shotgun sequencing of sorted derivative chromosomes is another way to map chromosomal breakpoints, but this approach is prohibitively costly if based on conventional Sanger sequencing. However, the recent introduction of novel multiplex sequencing-by-synthesis technology has reduced the costs of sequencing by several orders of magnitude. With the Solexa/Illumina system, which is based on massively parallel sequencing of millions of fragments using the proprietary Clonal Single Molecule Array technology and novel reversible terminator-based sequencing chemistry (for further details, see http://www.illumina.com/pages.ilmn?ID=203), >30 million short sequence reads can be produced in one single experiment. In theory, the density of this coverage should suffice to define primers for the amplification and subsequent Sanger sequencing of breakpoint-spanning PCR fragments.

Here, we have made use of this principle to map and sequence the breakpoint regions on flow-sorted derivative chromosomes from three unrelated mentally retarded patients with *de novo* balanced chromosomal translocations. In all of these patients, we identified suggestive novel candidate genes for mental retardation (MR), which were truncated by the chromosomal rearrangements. This study demonstrates the potential of this approach for large-scale breakpoint mapping and gene finding in patients with disease-associated balanced translocations.

Results

One derivative chromosome from each patient was flow-sorted and sequenced using the Solexa/Illumina 1 G analyzer, and sequence reads were aligned to reference sequences of the two corresponding chromosomes. Due to incomplete separation of derivative chromosomes from their normal counterparts upon flow sorting, we had to assess the transition point between high and low sequencing coverage by maximum likelihood estimation (see

Supplementary methods). This procedure, and the high number of sequencing reads generated by the 1G analyzer, enabled us to map all breakpoints precisely enough to allow direct amplification and sequencing of the junction fragments.

In patient 1 (46XY, t(7;9)(q36;q13), see Fig. 1A), the der (9) chromosome was sequenced in one lane of a flow cell. In total, 199421 and 1047649 reads mapped uniquely to chromosomes 7 and 9, respectively (Table 1). The average coverage of der (9) was approximately 18 reads per kb. Based on the distribution of sequence reads (Fig. 2), maximum likelihood estimation assigned the chromosome 7 breakpoint to position 156288272 and the breakpoint on chromosome 9 to position 71021542. To amplify the junction fragment of the der (9) chromosome, primers flanking the breakpoint were then designed and used for PCR amplification from patient genomic DNA. This yielded a PCR product of ~2.4 kb which was subsequently sequenced. Precise breakpoint junctions were determined through alignment with the reference sequences of chromosomes 7 and 9, and results were confirmed by PCR with primer pairs flanking the breakpoint on the der (7) chromosome. In this way, the respective breakpoints could be mapped between nucleotides 156288232 and 156288233 on chromosome 7, and between nucleotides 71021561 and 71021562 on chromosome 9 (NCBI build 35, Figure 4A).

The same procedure was applied to patient 2 (46,XY,t(4;5)(q21.1;p15.33), Fig. 1B) and patient 3 (46,XX,t(2;12)(q31;q22), Fig. 1C). In patient 2, the der (5) chromosome was sequenced. In total, 509023 and 641701 reads were obtained for chromosomes 4 and 5, respectively. The average coverage of der (5) was 3.7 reads per kb, which enabled us to localize the corresponding breakpoints at or near position 66330277 on chromosome 4 and close to position 2978264 on chromosome 5. By sequencing junction fragments, we could map the breakpoint on chromosome 4 to a deleted interval of 18 bps between nucleotides

66329510 and 66329529, and the breakpoint on chromosome 5 between base pairs 2977374 and 2977375 (NCBI build 35, see Fig. 4B). In patient 3, we sequenced the der (12). 201627 and 165160 reads mapped to unique sequences of chromosomes 2 and 12, respectively, with a mean coverage of 1.7 reads per kb for the der (12). Primers were designed on both sides of the estimated breakpoint position to amplify the junction fragments (Fig. 3). Subsequent Sanger sequencing mapped the breakpoint on chromosome 2 between nucleotides 84970997 and 84970999 with a 1 bp deletion, and the other breakpoint to a deleted interval of 407 bps between nucleotides 156795885 and 156796293 on chromosome 12 (NCBI build 35, figure 4C). Interestingly, 9 bp of the breakpoint sequence on the der (12) chromosome could neither be aligned to chromosome 12 nor to chromosome 2.

To identify interspersed repetitive elements in the vicinity of the breakpoints, we referred to the UCSC Genome Browser (NCBI build 35). In patient 1, the breakpoints on chromosomes 7 and 9 were spanned by DNA repeats of the MER1 and MER2 type, respectively. In patient 2, no repeats were found 1 kb upstream or downstream of the breakpoint on chromosome 4, whereas the breakpoint on chromosome 5 was within a short interspersed element (SINE) belonging to the Alu family. In patient 3, a DNA repeat of the MER1 type was detected within the breakpoint interval on chromosome 2, and a MaLR family repeat was found to span the breakpoint on chromosome 12.

Discussion

Balanced chromosome rearrangements can cause genetic diseases by disrupting or inactivating specific genes, forming a visible bridge between human phenotypes and genotypes. Characterisation of breakpoints in disease-associated balanced translocations (DBCRs) has led to the molecular elucidation of many hereditary diseases (Bugge et al. 2000; Tommerup 1993; Wirth et al. 1999). More than a decade ago, the Mendelian Cytogenetics

Network (MCN) has been established to study DBCRs in a systematic fashion (Bugge et al. 2000; Tommerup 1993), and more recently, similar programs to characterize DBCRs have been initiated elsewhere (Alkuraya et al. 2006; Lu et al. 2007; Williamson et al. 2007).

To speed up the characterization of breakpoints in patients with DBCRs and to overcome the resolution limits of conventional FISH mapping with BAC probes, we have now employed next generation sequencing as a fast and highly accurate way to localize the breakpoints in DNA from sorted derivative chromosomes. Solexa sequencing enabled us to map the chromosomal breakpoints with an error margin of less than 1000 bp, precisely enough to define a single primer pair for PCR amplification of the junction fragment, and to determine the exact sequence around the breakpoint in one single step. To infer the position of breakpoints, only sequencing reads mapping to unique sites in the genome were employed. Mapping of breakpoints may be very difficult with this procedure if they are located in large duplicated segments of the genome, but this is also true for other approaches like array painting.

As shown here, sequencing the breakpoint regions in three mentally retarded patients with reciprocal translocations has enabled us to identify three promising, truncated candidate genes, which had not been implicated in MR previously. In patient 1, the breakpoint on chromosome 7q36 maps to an intergenic region which is flanked by the 3' ends of two genes, *NOM1* and *MNX*; neither of these is a plausible functional candidate gene for MR. In contrast, the breakpoint on chromosome 9 disrupts 5' end of a splicing isoform (accession number: AL136545) of the *TRPM3* gene (Fig 5A). The product of this gene belongs to the family of transient receptor potential (TRP) channels, which are important for cellular calcium signalling and homeostasis. Given the expression of this isoform in brain, it is quite possible

that its disruption by the chromosome rearrangement contributes to the MR phenotype of patient 1.

In patient 2, the breakpoint on chromosome 4 disrupts *EPHA5*, a member of the Ephrin receptor family (figure 5B); there are no known genes in the vicinity of the other breakpoint on chromosome 5. EPH and EPH-related receptors have an important role in the establishment of topographic maps, especially in the nervous system (Drescher et al. 1997).. In *EPHA5* knockout mice, the retino-tectal map was abnormal, with temporal and nasal axons terminating in incorrect regions (Feldheim et al. 2004). It is therefore plausible that the clinical symptoms in patient 2 are related to the disruption of *EPHA5*.

Similarly, only one of the breakpoints of patient 3, that on chromosome 12, disrupts a known gene, and as in patient 1 and patient 2, it is an interesting candidate, which has not been directly implicated in MR yet. The relevant gene, *MGAT4C* (Figure 5C), encodes mannosyl (alpha-1,3-)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase, isozyme C, which is one of the key glycosyltransferases regulating the formation of tri- and other multiantennary Asn-linked sugar chains and expresses in fetal and adult brain (Gleeson and Schachter 1983; Schachter et al. 1989). Genetic defects of protein glycosyltransferases have emerged as an important cause of (mostly syndromic) forms of MR , very recently, mutations involving a putative member of this family, *TUSC3*, have been observed in two unrelated families with non-syndromic MR (F. Molinari 2007; Garshasbi et al. 2008;). Thus, it is quite conceivable that the MR in patient 3 is due to the truncation of the *MAGT4C* gene.

Due to the limited number of balanced translocations that have been analysed in detail, relatively little is still known about the involvement of specific sequences in the generation of such rearrangements. In the three reciprocal translocations described here, interspersed

repeats were detected at five out of six breakpoints. Repetitive elements have been implicated as the sites of chromosome instability, and the apparent enrichment of repetitive sequences around the breakpoints in this study may indicate that these elements have a role in the generation of balanced translocations. However, no significantly homologous sequences were found to flank the two breakpoints in each of the three patients. This argues against non-allelic homologous recombination (NAHR) as the mechanism underlying these rearrangements. Thus, it is more likely that they were generated through non-homologous end joining, which is in keeping with previous observations (Erdogan et al. 2006).

Hitherto, breakpoint analysis in patients with balanced chromosome rearrangements has focused on congenital or early-onset disorders. However, balanced chromosome rearrangements were also shown to be associated with complex and late-onset disorders, as documented by the identification of candidate genes for diabetes, schizophrenia, dyslexia, Tourette syndrome and psoriasis (Abelson et al. 2005; Blackwood et al. 2001; Gloyn et al. 2002; Taipale et al. 2003; Tzschach et al. 2006), and systematic re-examination of previously healthy carriers of balanced reciprocal translocations has identified several rearrangements that seem to predispose to late onset, complex diseases (Bache et al. 2006). However, given the frequency of complex disorders, it will be a major challenge to distinguish these 'causative' associations from the many others where apparent associations between balanced rearrangements and common disorders are due to chance. Therefore, our method, which greatly facilitates breakpoint mapping and sequencing should also be an asset for the identification of genetic factors in the etiology of common disorders.

In this study, we have used flow sorting to separate the derivative chromosomes from each other and from their normal counterparts, but depending on the size of the relevant chromosomes, this approach is not always possible, and it cannot be employed to study

inversions. As an alternative, microdissection has been used to isolate rearranged chromosomes, or specific chromosome region in case of inversions, and in principle it is feasible to use this approach in combination with next generation sequencing if chromosome sorting is no option. Finally, as the costs for high throughput sequencing will continue to decrease, whole genome paired-end sequencing may soon become a viable alternative to the procedure described here, allowing the characterization of DBCRs and the identification of the truncated genes in a single experiment, obviating prior isolation of rearranged chromosomes and chromosome segments. Technically, this is already feasible now (Korbel et al. 2007).

Materials and Methods

Patients

Patient 1

This male patient is the second child of healthy and non-consanguineous parents. He suffered from delayed psychomotor development and congenital bilateral cataracts that had been operated at the age of 6 and 16 months. On examination at the age of 17 years, he had severe visual defects and mild mental retardation. Body measurements were within the normal range. Cytogenetic investigation revealed a 46XY,t(7;9)(q36;q13) karyotype in the patient. Both parents had normal karyotypes.

Patient 2

This boy is the first child of healthy and non-consanguineous parents. Psychomotor development was retarded (walking at age 2 years). On examination at the age of 8 years, he was severely mentally retarded. He was not able to talk and walked with an insecure gait. Brain MR scan showed no abnormalities. He had a horizontal nystagmus, but his vision was apparently normal and ophthalmological investigations revealed no structural abnormalities.

Cytogenetic investigation revealed a 46,XY,t(4;5)(q21.1;p15.33) karyotype in the patient. Karyotypes were normal in his two parents.

Patient 3

This female patient is the first child of healthy and non-consanguineous parents. Her psychomotor development was delayed; at the age of 17 months she could neither walk nor talk. An MRI scan revealed frontotemporal atrophy and delayed myelinisation. Body measurements were within the normal range. She also suffered from sleep apnoea. Cytogenetic investigations revealed a 46,XX,t(2;12)(q31;q22) karyotype in the patient and normal karyotypes in her parents.

Exclusion of clinically relevant DNA copy number variation

To study submicroscopic deletions or duplications, comparative genome hybridisation (CGH) experiments were performed in all three patients, using whole genome tiling path BAC arrays, as previously described (Erdogan et al. 2006), and results were compared with previously described copy number variants (CNVs) and CNVs observed in ~ 700 unrelated probands screened for genomic imbalances in our laboratory using tiling path BAC arrays (Kirov et al. 2007; Ullmann et al. 2007). Except for a deletion of ~360kb on chromosome 5 (position 101,033,643-101,393,788 NCBI build 35) in patient 1, no genomic imbalances larger than three consecutive clones were found that had not been observed also in healthy individuals. The 360 kb deletion on chromosome 5q encompasses two known CNVs (Locus 1513 and 1514, Database of Genome Variants, http://projects.tcag.ca/variation/), does not span known genes, and may be functionally neutral. Familial inheritance could not be studied due to lack of parental DNA. Array CGH data discussed in this publication have been deposited in NCBIs Gene Expression Omnibus (Barrett et al. 2007) and are accessible through GEO Series accession number GSE10115.

Chromosome sorting and amplification

For chromosome sorting, the lymphoblastoid cell lines were cultured in RPMI 1640 medium supplemented with 10% fetal calf serum, 2 mM L-glutamine, and antibiotics at 37°C in a humidified atmosphere containing 5% CO2. Cells in log phase were treated for 16h with colcemid (0.05 mg/mL final concentration) to arrest cells in metaphase. Metaphase chromosomes were flow sorted as described previously (Arkesteijn et al. 1999) . The sorted chromosomes were amplified using the GenomiPhi V2 DNA Amplification Kit following the protocol of the manufacturer.

Chromosome sequencing using Solexa

Approximately 2 ug amplified chromosomes were randomly fragmented to less than 800bp by nebulisation. Both ends of DNA fragments were then repaired by T4 polymerase and Klenow DNA polymerase, and T4 polynucleotide kinase. Using Klenow exo (3' to 5' exo minus), a protruding 'A' base was added to 3' end of DNA fragments for ligation with Solexa adaptors (with a 'T' overhang). Following ligation, DNA fragments (size range 150 to 200bp) were gel purified on 2% agarose, followed by 18 cycles of PCR-amplification. We measured the DNA concentration with a Nanodrop 7500 spectrophotometer, and a 1 ul aliquot was diluted to 10nM. Adaptor-ligated DNA was hybridized to the surface of flow cells, and DNA clusters were generated using the Illumina/Solexa cluster station, followed by 27 to 36 cycles of sequencing on the Illumina/Solexa 1G analyzer, in accordance with the manufacturer's protocols.

Data Processing

Sequence reads were compiled using a manufacturer-provided computational pipeline consisting of the open source Firecrest and Bustard applications. Sequence reads from a

derivative chromosome were then aligned with the sequence of its 2 normal counterparts (NCBI build 35) using the Eland application. Only uniquely mapped reads with less than 2 mismatches were retained. We then calculated the number of reads overlapping each nucleotide position along a specific chromosome. The change point problem is a natural framework to approach the breakpoint estimation problem. We used the Poisson process to model the random generation of sequencing reads along a chromosome. The breakpoint position was computed using the method of maximum likelihood estimation. Computer simulations confirmed the robustness of these estimates (for a more detailed description, see Supplementary Methods).

PCR amplification and sequencing of junction fragments

Junction fragments were amplified by long range PCR using the Takara LA PCR kit version 2.1. Primers are listed in Supplementary Table, and PCR conditions are available upon request. PCR products were used as templates for sequencing in both directions using BigDye Terminator chemistry (PE Biosystems) on an Applied Biosystem 3730xl DNA Analyzer. Sequences of junction fragments were aligned to the human genome reference sequence (NCBI build 35) using Blat from the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway).

Acknowledgements

The derivative chromosomes of patient 1 were sorted by J. Wienberg, Chrombios GmbH, Raubling, Germany. The project was supported by the German National Genome Research Network (project no.s 01GR0105 and 01GR0414) and by a grant from the Max-Planck Innovation Funds (to HHR).

Legends to Figures

Figure 1: Ideograms of the derivative chromosomes from the patients. A. patient 1. B. patient 2. C. patient 3. Breakpoints are marked by arrows.

Figure 2: Solexa sequencing profile of derivative chromosome 9 from patient 1. 1 Mb intervals around the breakpoints on chromosome 7 (A) and 9 (B) are shown. 199,421 and 1,047,649 reads derived from der(9) were mapped to unique positions on normal chromosome 7 and 9, respectively. The number of reads was then binned into non-overlapping 1kb segments and plotted against the chromosome coordinates. Breakpoints are marked by arrows.

Figure 3: PCR products of the 2 junction fragments from patient 3. M. size marker C. genomic DNA from a normal individual. P. genomic DNA from patient 3. Der, derivative chromosome.

Figure 4: Junction fragment sequences in patients 1 (A), 2 (B) and 3 (C). Normal reference sequences are labelled in Italic and normal characters, respectively. Deleted sequences are underlined. The 9 bp insertion on der (12) of patient 3 is marked by a black box. Chr, chromosome; Der, derivative chromosome.

Figure 5: Chromosome breakpoints (marked by arrows) and disrupted genes

A. Patient 1: breakpoint on chromosome 7 maps between the 3' ends of *NOM1* and *MNX1*; breakpoint on chromosome 9 disrupts the 5' end of a splicing isoform (accession number: AL136545) of *TRPM3*. B. Patient 2: breakpoint on chromosome 4 disrupts *EPHA5*; no known genes in the vicinity of the breakpoint on chromosome 5. C. Patient 3: breakpoint on chromosome 12 disrupts *MAGT4C*; no known genes close to the breakpoint on chromosome 2.

Figures

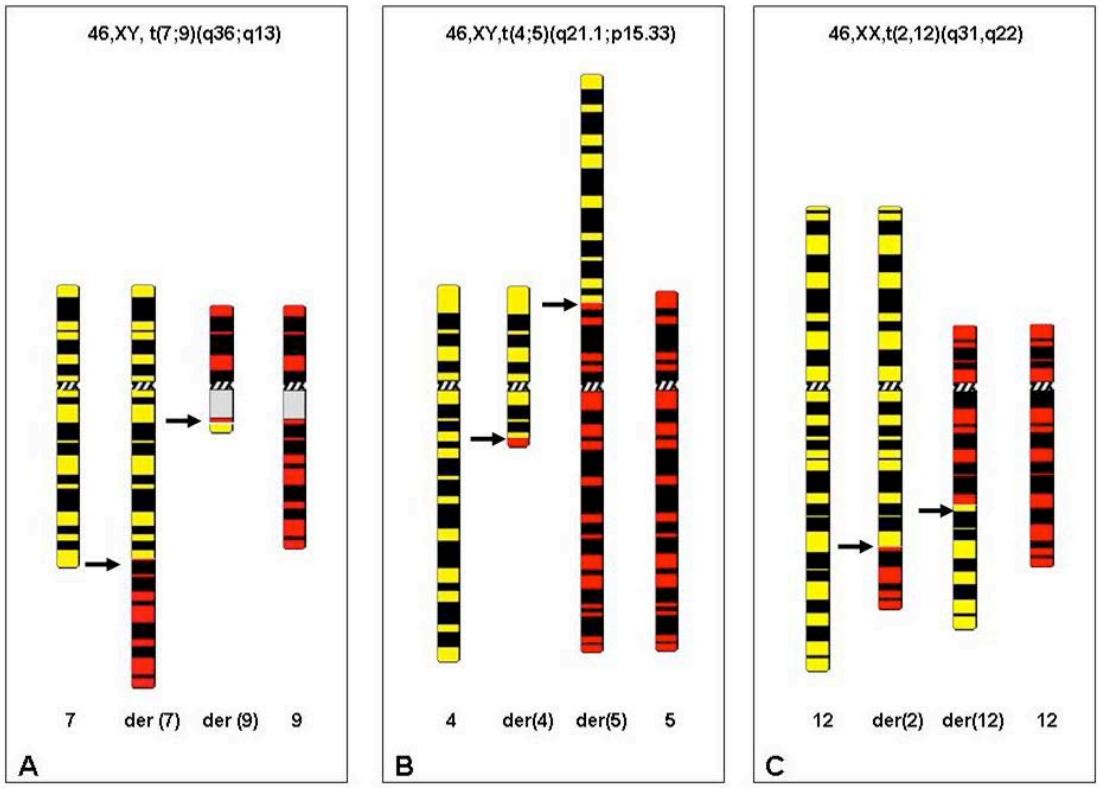

Figure 1

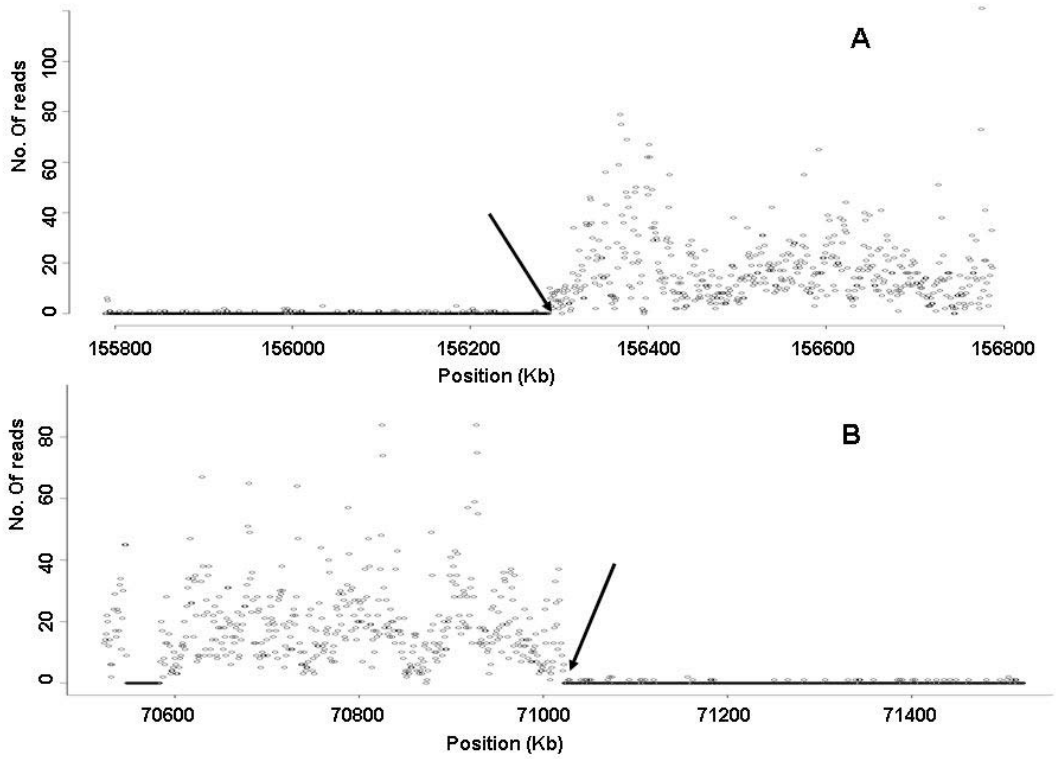

Figure 2

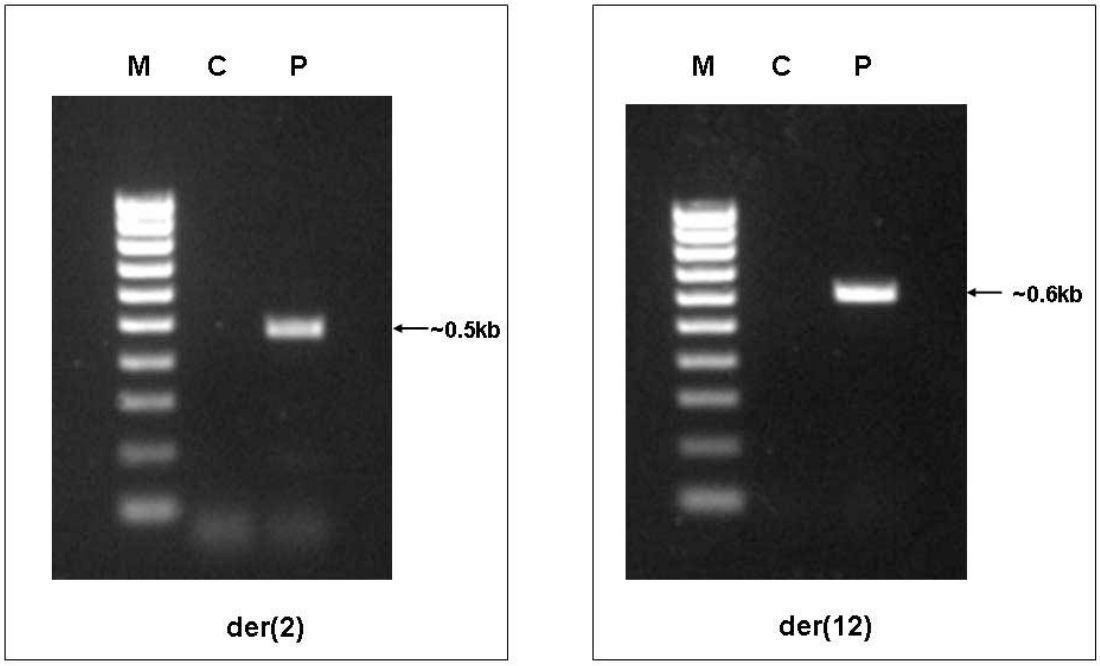

Figure 3

A

Chr7 (+) *CCCCTGCCCCTCCCCCTCCCCTCCATTCCCCTCCCTTCCTTCTTTTTTCTTTTGTTAGAT*

Der7 *CCCCTGCCCCTCCCCCTCCCCTCCATTCCCC*AGGAGCATGAACCCTGCTGTGAACTGTGC

Der9 TGTCACATTAGCATTGGCATTAAATTCTCA*CTCCCTTCCTTCTTTTTTCTTTTGTTAGAT*

Chr9 (+) TGTCACATTAGCATTGGCATTAAATTCTCACAGGAGCATGAACCCTGCTGTGAACTGTGC

B

Chr4 (+) *ACTATTGTCAAAATCTGAGTACCAGAATGGAAAATCCAGCTTAAGTTCTTCTTGATTCAG*

Der4 *ACTATTGTCAAAATCTGAGTACCAGAATGC*AACTCAGGAGGTGGAGGTTGCAGTGAGCCG

Der5 TACTGGGGAGGCTGAGGCAGGAGAATCAC *CTTCTTGATTCAG*

Chr5 (-) TACTGGGGAGGCTGAGGCAGGAGAATCACCAACTCAGGAGGTGGAGGTTGCAGTGAGCCG

C

Chr2 (+) *CAATGAAGTTAGAATCTCAAATAATGGAG...CTGGTCACTTTGTAACAAGCAAAAGCA*

Der2 *CAATGAAGTTAGAATCTCAAATAATG* TGGTCTCAAAGAAGAGAGAATTGAAGG

Der12 CACAAGGTGAAAGGCACGTCCTAC TTTGTACTT *GTCACTTTGTAACAAGCAAAAGCA*

Chr12(+) CACAAGGTGAAAGGCACGTCCTACATGGTCTCAAAGAAGAGAGAATTGAAGG

Figure 4

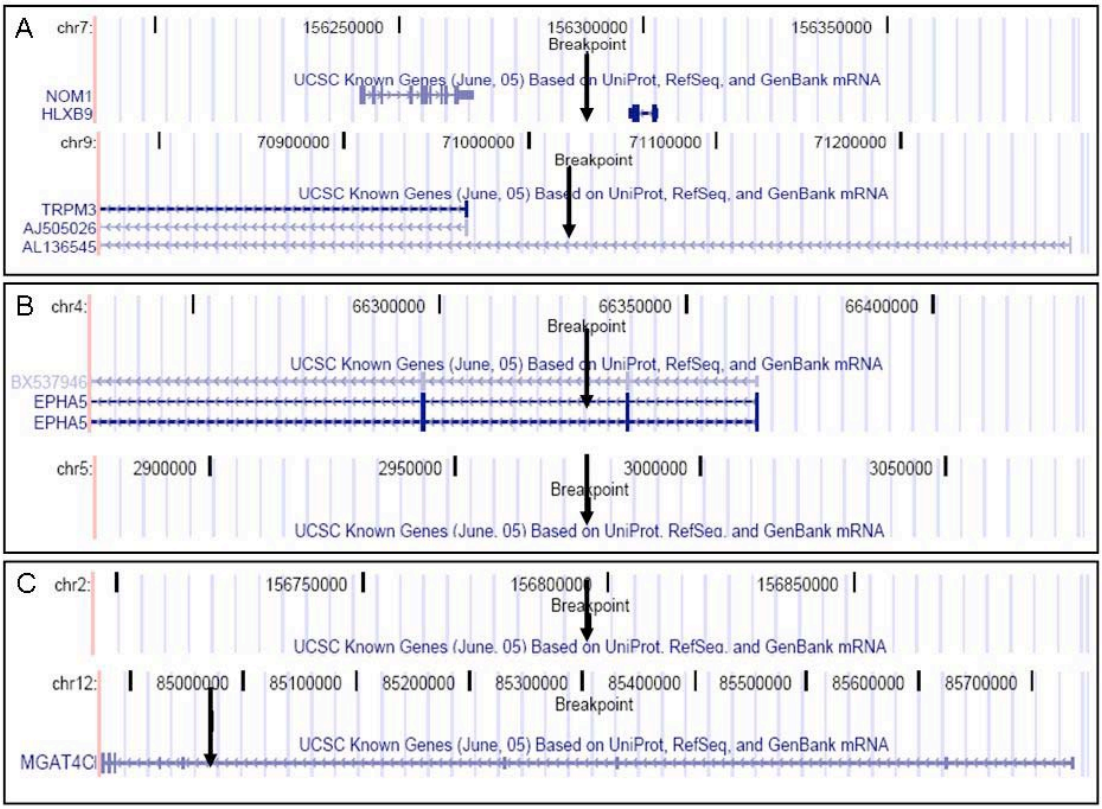

Figure 5

Table

Table 1: Breakpoint position estimated from Solexa sequencing and direct Sanger sequencing.

Patient	Chromosome	No. of mapped Solexa sequencing reads	Breakpoints estimated from Solexa sequencing	Breakpoint position from Sanger sequencing
1	7	199,421	156288272	156288232-156288233
	9	1,047,649	71021542	71021561-71021562
2	4	509,023	66330277	66329510-66329529
	5	641,701	2978264	2977374-2977375
3	2	201,627	156796485	156795885-156796293
	12	165,160	84970972	84970997-84970999

References

Abelson, J.F., K.Y. Kwan, B.J. O'Roak, D.Y. Baek, A.A. Stillman, T.M. Morgan, C.A. Mathews, D.L. Pauls, M.R. Rasin, M. Gunel, N.R. Davis, A.G. Ercan-Sencicek, D.H. Guez, J.A. Spertus, J.F. Leckman, L.S.t. Dure, R. Kurlan, H.S. Singer, D.L. Gilbert, A. Farhi, A. Louvi, R.P. Lifton, N. Sestan, and M.W. State. 2005. Sequence variants in SLITRK1 are associated with Tourette's syndrome. *Science* **310:** 317-320.

Alkuraya, F.S., I. Saadi, J.J. Lund, A. Turbe-Doan, C.C. Morton, and R.L. Maas. 2006. SUMO1 haploinsufficiency leads to cleft lip and palate. *Science* **313:** 1751.

Arkesteijn, G., E. Jumelet, A. Hagenbeek, E. Smit, R. Slater, and A. Martens. 1999. Reverse chromosome painting for the identification of marker chromosomes and complex translocations in leukemia. *Cytometry* **35:** 117-124.

Bache, I., M. Hjorth, M. Bugge, S. Holstebroe, J. Hilden, L. Schmidt, K. Brondum-Nielsen, G. Bruun-Petersen, P.K. Jensen, C. Lundsteen, E. Niebuhr, K. Rasmussen, and N. Tommerup. 2006. Systematic re-examination of carriers of balanced reciprocal translocations: a strategy to search for candidate regions for common and complex diseases. *Eur J Hum Genet* **14:** 410-417.

Backx, L., H. Van Esch, C. Melotte, N. Kosyakova, H. Starke, J.P. Frijns, T. Liehr, and J.R. Vermeesch. 2007. Array painting using microdissected chromosomes to map chromosomal breakpoints. *Cytogenet Genome Res* **116:** 158-166.

Barrett, T., D.B. Troup, S.E. Wilhite, P. Ledoux, D. Rudnev, C. Evangelista, I.F. Kim, A. Soboleva, M. Tomashevsky, and R. Edgar. 2007. NCBI GEO: mining tens of millions of expression profiles--database and tools update. *Nucleic Acids Res* **35:** D760-765.

Blackwood, D.H., A. Fordyce, M.T. Walker, D.M. St Clair, D.J. Porteous, and W.J. Muir. 2001. Schizophrenia and affective disorders--cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family. *Am J Hum Genet* **69:** 428-433.

Bugge, M., G. Bruun-Petersen, K. Brondum-Nielsen, U. Friedrich, J. Hansen, G. Jensen, P.K. Jensen, U. Kristoffersson, C. Lundsteen, E. Niebuhr, K.R. Rasmussen, K. Rasmussen, and N. Tommerup. 2000. Disease associated balanced chromosome rearrangements: a resource for large scale genotype-phenotype delineation in man. *J Med Genet* **37:** 858-865.

Drescher, U., F. Bonhoeffer, and B.K. Muller. 1997. The Eph family in retinal axon guidance. *Curr Opin Neurobiol* **7:** 75-80.

Erdogan, F., W. Chen, M. Kirchhoff, V.M. Kalscheuer, C. Hultschig, I. Muller, R. Schulz, C. Menzel, T. Bryndorf, H.H. Ropers, and R. Ullmann. 2006. Impact of low copy repeats on the generation of balanced and unbalanced chromosomal aberrations in mental retardation. *Cytogenet Genome Res* **115:** 247-253.

F. Molinari, F.F., W. Morelle, S. Romano, P. de Lonlay, P.S. Tarpey, J. Teague, S. Edkins, P.A. Futreal, M.R. Stratton, M. Partington, G. Matthijs, G. Turner, J. Gecz, A. Munnich, L. Colleaux. 2007. **Oligosaccharyltransferase subunits mutations in non-syndromic mental retardation**. In ***13th International Workshop on Fragile X and X-Linked Mental Retardation***, Venice, Italy.

Feldheim, D.A., M. Nakamoto, M. Osterfield, N.W. Gale, T.M. DeChiara, R. Rohatgi, G.D. Yancopoulos, and J.G. Flanagan. 2004. Loss-of-function analysis of EphA receptors in retinotectal mapping. *J Neurosci* **24:** 2542-2550.

Fiegler, H., S.M. Gribble, D.C. Burford, P. Carr, E. Prigmore, K.M. Porter, S. Clegg, J.A. Crolla, N.R. Dennis, P. Jacobs, and N.P. Carter. 2003. Array painting: a method for the rapid analysis of aberrant chromosomes using DNA microarrays. *J Med Genet* **40:** 664-670.

Garshasbi M., H.V., et al. 2008. A Defect in the TUSC3 Gene is Associated with Autosomal Recessive Mental Retardation. *Am J Hum Genet*. (In Press)

Gleeson, P.A. and H. Schachter. 1983. Control of glycoprotein synthesis. *J Biol Chem* **258:** 6162-6173.

Gloyn, A.L., S. Ellard, M. Shepherd, R.T. Howell, E.M. Parry, A. Jefferson, E.R. Levy, and A.T. Hattersley. 2002. Maturity-onset diabetes of the young caused by a balanced translocation where the 20q12 break point results in disruption upstream of the coding region of hepatocyte nuclear factor-4alpha (HNF4A) gene. *Diabetes* **51:** 2329-2333.

Gribble, S.M., D. Kalaitzopoulos, D.C. Burford, E. Prigmore, R.R. Selzer, B.L. Ng, N.S. Matthews, K.M. Porter, R. Curley, S.J. Lindsay, J. Baptista, T.A. Richmond, and N.P. Carter. 2007. Ultra-high resolution array painting facilitates breakpoint sequencing. *J Med Genet* **44:** 51-58.

Kirov, G., D. Gumus, W. Chen, N. Norton, L. Georgieva, M. Sari, M.C. O'Donovan, F. Erdogan, M.J. Owen, H.H. Ropers, and R. Ullmann. 2007. Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Hum Mol Genet*.

Korbel, J.O., A.E. Urban, J.P. Affourtit, B. Godwin, F. Grubert, J.F. Simons, P.M. Kim, D. Palejev, N.J. Carriero, L. Du, B.E. Taillon, Z. Chen, A. Tanzer, A.C. Saunders, J. Chi, F. Yang, N.P. Carter, M.E. Hurles, S.M. Weissman, T.T. Harkins, M.B. Gerstein, M. Egholm, and M. Snyder. 2007. Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318:** 420-426.

Lu, W., F. Quintero-Rivera, Y. Fan, F.S. Alkuraya, D.J. Donovan, Q. Xi, A. Turbe-Doan, Q.G. Li, C.G. Campbell, A.L. Shanske, E.H. Sherr, A. Ahmad, R. Peters, B. Rilliet, P. Parvex, A.G. Bassuk, D.J. Harris, H. Ferguson, C. Kelly, C.A. Walsh, R.M. Gronostajski, K. Devriendt, A. Higgins, A.H. Ligon, B.J. Quade, C.C. Morton, J.F. Gusella, and R.L. Maas. 2007. NFIA haploinsufficiency is associated with a CNS malformation syndrome and urinary tract defects. *PLoS Genet* **3:** e80.

Masoud Garshasbi, V.H., Haleh Habibi, Kimia Kahrizi, Roxana Kariminejad, Farkhondeh Behjati, Andreas Tzschach, Hossein Najmabadi, Hans Hilger Ropers, Andreas Walter Kuss. 2008. A Defect in the TUSC3 Gene is Associated with Autosomal Recessive Mental Retardation. *Am J Hum Genet*. (In Press)

Schachter, H., I. Brockhausen, and E. Hull. 1989. High-performance liquid chromatography assays for N-acetylglucosaminyltransferases involved in N- and O-glycan synthesis. *Methods Enzymol* **179:** 351-397.

Taipale, M., N. Kaminen, J. Nopola-Hemmi, T. Haltia, B. Myllyluoma, H. Lyytinen, K. Muller, M. Kaaranen, P.J. Lindsberg, K. Hannula-Jouppi, and J. Kere. 2003. A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. *Proc Natl Acad Sci U S A* **100:** 11553-11558.

Tommerup, N. 1993. Mendelian cytogenetics. Chromosome rearrangements associated with mendelian disorders. *J Med Genet* **30:** 713-727.

Tzschach, A., K. Hoffmann, M. Hoeltzenbein, I. Bache, N. Tommerup, C. Bommer, H. Korner, V. Kalscheuer, and H.H. Ropers. 2006. Molecular characterization of a balanced chromosome translocation in psoriasis vulgaris. *Clin Genet* **69:** 189-193.

Ullmann, R., G. Turner, M. Kirchhoff, W. Chen, B. Tonge, C. Rosenberg, M. Field, A.M. Vianna-Morgante, L. Christie, A.C. Krepischi-Santos, L. Banna, A.V. Brereton, A. Hill, A.M. Bisgaard, I. Muller, C. Hultschig, F. Erdogan, G. Wieczorek, and H.H. Ropers. 2007. Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. *Hum Mutat* **28:** 674-682.

Veltman, I.M., J.A. Veltman, G. Arkesteijn, I.M. Janssen, L.E. Vissers, P.J. de Jong, A.G. van Kessel, and E.F. Schoenmakers. 2003. Chromosomal breakpoint mapping by arrayCGH using flow-sorted chromosomes. *Biotechniques* **35:** 1066-1070.

Warburton, D. 1991. De novo balanced chromosome rearrangements and extra marker chromosomes identified at prenatal diagnosis: clinical significance and distribution of breakpoints. *Am J Hum Genet* **49:** 995-1013.

Williamson, R.E., K.N. Darrow, S. Michaud, J.S. Jacobs, M.C. Jones, D.F. Eberl, R.L. Maas, M.C. Liberman, and C.C. Morton. 2007. Methylthioadenosine phosphorylase (MTAP) in hearing: gene disruption by chromosomal rearrangement in a hearing impaired individual and model organism analysis. *Am J Med Genet A* **143:** 1630-1639.

Wirth, J., H.G. Nothwang, S. van der Maarel, C. Menzel, G. Borck, I. Lopez-Pajares, K. Brondum-Nielsen, N. Tommerup, M. Bugge, H.H. Ropers, and T. Haaf. 1999. Systematic characterisation of disease associated balanced chromosome rearrangements by FISH: cytogenetically and genetically anchored YACs identify microdeletions and candidate regions for mental retardation genes. *J Med Genet* **36:** 271-278.